



Australian Bureau of Statistics

6311.0 - Information Paper: Construction of Experimental Statistics on Employee Earnings and Jobs from Administrative Data, Australia, 2011-12

Latest ISSUE Released at 11:30 AM (CANBERRA TIME) 11/12/2015 First Issue

Summary

Executive Summary

EXECUTIVE SUMMARY

In 2015, the Australian Bureau of Statistics (ABS) assessed the feasibility of constructing experimental statistics on employee earnings and jobs from integrated person and business level administrative data from the Australian Taxation Office (ATO). In undertaking this assessment, the ABS complied fully with the High Level Principles for Data Integration Involving Commonwealth Data for Statistical and Research Purposes.

The purpose of the assessment was to:

- Demonstrate that person and business level taxation data can be integrated to generate experimental statistics that could fill a current gap in the available statistics or complement existing statistics.
- Demonstrate the utility of a future Linked Employer-Employee Database (LEED) for statistical purposes.

The Integrated Dataset used in this assessment was created by linking:

- person level data from the ATO Personal Income Tax (PIT) dataset, and
- business level taxation data from the ABS Expanded Analytical Business Longitudinal Database (EABLD). The EABLD incorporates the ABS Business Register (ABSBR) with business taxation data from the ATO and ABS business survey data.

Creating and using the Integrated Dataset has demonstrated that:

- integrating PIT and EABLD data is feasible;
- the Integrated Dataset can provide coherent information on employee earnings and jobs, as well as the number of employees in Australia; and
- further work is required to address identified limitations with the Integrated Dataset and the experimental statistics.

This paper provides the background to this assessment (hereafter called the LEED Foundation Projects), a description of the data sources, integration methodology and confidentiality process, summary of results and limitations of the experimental statistics. In addition, the paper outlines directions towards a future LEED. A data cube of the aggregate experimental statistics on employee earnings and jobs is available from the Downloads tab of this release.

The ABS is seeking feedback on the LEED Foundation Projects and the creation of an enduring LEED through this release.

Introduction

INTRODUCTION

The paper summarises the approach taken to construct the Integrated Dataset (linked employer-employee data) and experimental statistics on employee earnings and jobs undertaken as part of the Linked Employer-Employee Database (LEED) Foundation Projects. The paper provides information on the data sources and integration methodology, a summary of results, and future directions for the LEED in the ABS.

The paper is structured as follows:

- Introduction - overview of the LEED Foundation Projects, their scope and coverage;
- Data Sources - Personal Income Tax (PIT) and the Expanded Analytical Business Longitudinal Database (EABLD);
- Integration Methodology - construction of the Integrated Dataset;
- Summary of Results - aggregate experimental statistics and their coherence with ABS estimates; and
- Future Directions - microdata product and future LEED.

The Abbreviations, [Glossary](#), and [appendices](#) are provided in the [Explanatory Notes](#).

CONTEXT

The Australian Bureau of Statistics (ABS) is embarking on a period of major organisational transformation to respond to the new opportunities and challenges of the dynamic statistical landscape. Maximising the value of administrative data through integration and improved access is a strategic priority for the ABS in order to deliver high quality official statistics in efficient and innovative ways. For more information, refer to the [ABS Corporate Plan, 2015-19](#) (cat. no. 1005.0).

As part of this transformation, the ABS is exploring the potential of creating a LEED which integrates person and business administrative data sourced from the Australian Taxation Office (ATO). The LEED would be linked longitudinally, as well as provide point in time data.

A future LEED would build on the EABLD, which integrates business tax and survey data. The LEED would be created by integrating the EABLD with a longitudinally linked PIT database. The long term vision is to extend the LEED by integrating other key administrative data, survey data (person and business level), and data from the Census of Population and Housing to deliver a new statistical solution to vastly expand the information base on the Australian labour market.

The LEED would address a longstanding information gap in Australian labour statistics by being a single database capable of addressing complex and varied questions about employer-employee relationships at both a point in time and longitudinally (e.g. examining firm and employee characteristics of productive firms). The creation of a LEED would demonstrate that administrative and directly collected data can be integrated to provide a strong evidence base for research, policy development and evaluation.

LEED FOUNDATION PROJECTS

The LEED Foundation Projects are being undertaken to build support for the future LEED. The purpose of these projects is to demonstrate the value of the LEED by assessing the feasibility of integrating person and business tax data, and using this Integrated Dataset to create new statistical outputs.

The LEED Foundation Projects integrate person level data from the PIT dataset with business level data sourced from the EABLD (integrated business data) to produce experimental statistics on employee earnings and jobs for the 2011-12 financial year.

These projects demonstrate an important step towards a future LEED, as it is the first time the ABS has integrated PIT data with the EABLD.

The experimental statistics produced from the Integrated Dataset (linked employer-employee data) were designed to:

- assess whether administrative data can be used to:
 - address a known information gap by creating a new experimental measure of employee jobs in Australia; and
 - produce new experimental statistics to complement existing information on employees and earnings from ABS household and business collections;
- be an example of the value of a future LEED for statistical purposes.

SCOPE

The LEED Foundation Projects capture information on all employee earnings and jobs in Australia throughout the reference period of 1 July 2011 to 30 June 2012.

The scope of the LEED Foundation Projects includes:

- all persons who were an employee (see below) at any point in the reference period as recorded on either an Individual Tax Return (ITR) or an Individual Pay As You Go (PAYG) summary;
- all jobs as reported in an Individual PAYG summary during the reference period; and
- all businesses which provided an Individual PAYG summary to an employee in the reference period.

An **employee** (see [Explanatory Notes](#), paragraphs 32-35) is defined as someone who reported earnings on an ITR (see [Explanatory Notes](#), paragraphs 36-51) or who had an Individual PAYG summary reporting \$1 or more in gross payment (see [Explanatory Notes](#), paragraph 61).

Persons who did not report any earnings on their ITR and did not receive an Individual PAYG summary from an employer were excluded from the scope of the LEED Foundation Projects. These include persons who were not in the labour force, were unemployed, or were employed but were away from all of their jobs for the entire reference period and did not receive any pay during that period.

A **job** is defined as a link between an employee and a business for \$1 or more in payment as reported on an Individual PAYG summary. An employee can have multiple jobs with the same or different businesses during the financial year, and can hold two or more jobs concurrently (see [Explanatory Notes](#), paragraphs 52-60).

Jobs in which no Individual PAYG summary was provided by the employing business are not captured in the PIT data, and are therefore not included. Jobs in which the occupier was an Owner Manager of an Unincorporated Enterprise (OMUE, e.g. sole traders) are out of scope as they are not considered an employee (although the person may be included as an employee in other jobs they may hold).

Businesses which did not report PAYG withholdings from employees are deemed to be non-employing and are out of scope of the LEED Foundation Projects. These businesses are deemed to be non-employing, irrespective of their employment size on the ABS Business Register, because they do not report any jobs through an Individual PAYG summary.

COVERAGE

Coverage restrictions apply to the LEED Foundation Projects Integrated Dataset.

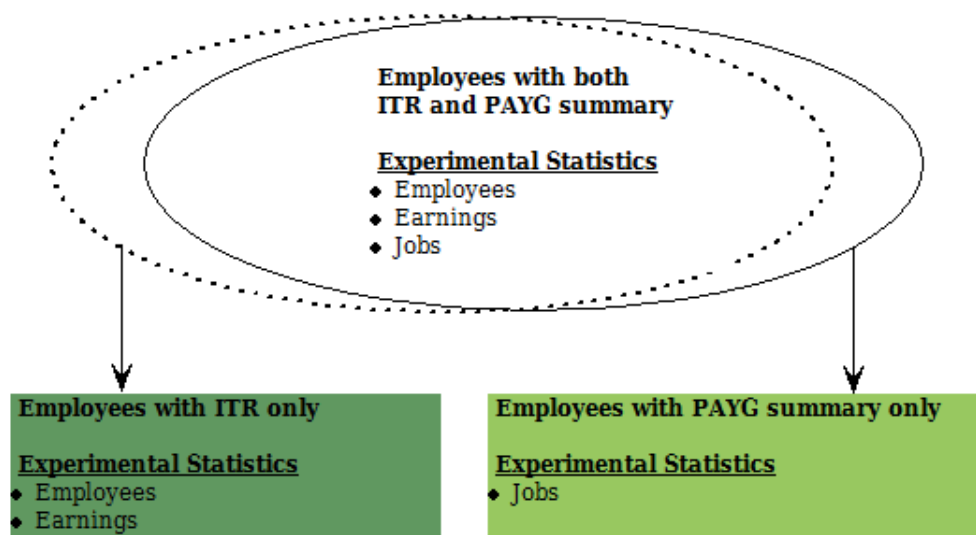
The LEED Foundation Projects' use of unique identifiers ensures that each individual is unlikely to be included more than once in the experimental statistics.

Employees who meet one of the following conditions will be partially excluded from the LEED Foundation Projects. For these employees, missing information from one source (e.g. missing PAYG data) will result in exclusion from certain statistics (e.g. Mean gross payment, or Number of jobs).

1. Employees who did not report earnings on an ITR for any of the following reasons:
 - They did not submit an ITR for any of the reasons outlined on pages 6 and 7 of the Individual Tax Return Instructions 2012 on the [ATO website](#);
 - They did not submit an ITR for any other reason; or
 - They submitted an ITR but did not report any earnings.
2. Employees who did not receive an Individual PAYG summary from an employer for any of the following reasons:
 - They worked for cash in hand or other payments not recorded on an Individual PAYG summary;
 - They conducted illicit activities not recorded on Individual PAYG summaries;
 - They did not supply their Tax File Number (TFN) to their employer; or
 - Any other reason.

No employing businesses were excluded on the basis of coverage.

Diagram 1: Implications of coverage on experimental statistics on employee earnings and jobs



CONFIDENTIALITY

The results of the LEED Foundation Projects are based, in part, on:

- Taxation data supplied by the ATO to the Australian Statistician under the **Taxation Administration Act 1953**; and
- Australian Business Register (ABR) data supplied by the Registrar to the Australian Statistician under **A New Tax System (Australian Business Number) Act 1999**.

These Acts require that such data is only used by the ABS for the purpose of administering the **Census and Statistics Act 1905**. The ABS is obligated to maintain the confidentiality of

individuals and businesses in these ATO and ABR data sets, as well as comply with provisions that govern the use and release of this information, including the **Privacy Act 1988**.

Access to taxation data is tightly controlled within the ABS. Policies and guidelines governing the disclosure of information were implemented and followed in order to maintain the confidentiality of individuals and businesses. The aggregate experimental statistics have been confidentialised to ensure that they are not likely to enable identification of a particular person or organisation.

Data Sources

DATA SOURCES

The LEED Foundation Projects integrated person and business level data for the 2011-2012 financial year to construct the Integrated Dataset. The two data sources are the PIT data and the EABLD.

Any discussion of limitations or weaknesses in relation to ATO or ABR data is in the context of using the information for statistical purposes, and is not related to the ability of the data to support the ATO or ABR's core operational requirements.

PERSONAL INCOME TAX (PIT) DATASET

The PIT dataset is person level unit record data compiled by the ATO. It is provided to the ABS in three subsets, and the data for each individual can be linked using an encrypted person identifier, the Scrambled Tax File Number (STFN). For the purpose of the LEED Foundation Projects, select variables were extracted from each of the subsets, as described below.

Client Register

This contains demographic information for each person who has submitted an ITR (or had some other transaction with the ATO) at some point in time. This is a constantly evolving register which is updated using information from various sources such as an ITR. As a result, the information is referenced to the date at which extracts are taken by the ATO and provided to the ABS. The reference date for the Client Register file for the LEED Foundation Projects is July 2014.

The following key variables are sourced from the Client Register:

- Scrambled TFN (STFN) of employee
- Geocoded address (State and territory and Statistical Area Level 4).

Client Dataset

This contains personal income information for each person who lodged an ITR with the ATO. This data is aggregated to the person level and contains information such as Earnings (see Explanatory Notes, paragraphs 36-51) and Occupation in main job (see Explanatory Notes, paragraphs 62-63). Data provided to the ABS by the ATO are from taxation returns processed up to 16 months after the end of the financial year (i.e. returns processed up to 31 October 2013 for the financial year ending 30 June 2012).

The following key variables were extracted from the Client Dataset:

- STFN of employee
- Sex
- Age
- Occupation in main job

- Salary or wages
- Allowances, earnings, tips, directors fees etc.
- Employer lump sum payments
- Attributed personal services income
- Employee share schemes, total assessable discount amount
- Total reportable fringe benefits amounts
- Reportable employer superannuation contributions.

For further information on the ITR data items referred to above, please refer to the Individual Tax Return Instructions 2012 on the ATO website.

Individual Pay As You Go (PAYG) Dataset

This contains job level information reported by employers about gross payments made to employees and the start and end dates for each job. This dataset contains both the STFAN of the employee and the Australian Business Number (ABN) of the employer.

The following key variables were extracted from the Individual PAYG Dataset:

- STFAN of employee
- ABN of employer
- Gross payment amounts
- Employment date information.

EXPANDED ANALYTICAL BUSINESS LONGITUDINAL DATABASE (EABLD)

The business unit record data used in the LEED Foundation Projects comes from the EABLD, construction of which was completed in 2015. The EABLD is based on the ABS Business Register (ABSBR), which uses the ABS Units Model to describe the characteristics of businesses and the structural relationships between related businesses.

For further information on the construction of the EABLD, refer to Information Paper: Construction of the Expanded Analytical Business Longitudinal Database, 2001-02 to 2012-13 (cat. no. 8171.0).

For further information on the ABS Units Model, refer to the Appendix of the Standard Economic Sector Classifications of Australia, 2008 (cat. no. 1218.0).

The LEED Foundation Projects used an extract of the EABLD for the 2011-12 financial year containing selected variables (as described below). The EABLD extract contains all businesses registered up to and including the 2011-12 financial year. These are separated into two populations, as described below.

Non-profiled population (simple businesses)

The majority of businesses have simple structures and the unit registered for an ABN will satisfy ABS statistical reporting requirements. These businesses form the non-profiled population in the EABLD extract. The ABN is the statistical unit used in the LEED Foundation Projects to represent businesses in the non-profiled population.

Profiled population (complex businesses)

For those businesses where the ABN is not considered suitable for ABS statistical requirements, the ABS maintains its own units structure (the ABS Units Model) through direct contact with businesses. This population, known as the profiled population, consists typically of large, diverse and complex structured businesses. For businesses in the profiled population, statistical units include the Enterprise Group (EG) and the Type of Activity Unit (TAU). The range of activities carried out across the EG can be very diverse. The TAU is established to represent a grouping of one or more businesses within the EG that cover all the operations within an industry sub-division

and for which a basic set of financial, production and employment data can be reported. The TAU is the statistical unit used in the LEED Foundation Projects to represent EGs in the profiled population, such that each TAU of a complex business (EG) is considered to be a separate business.

The following key variables were extracted from the EABLD:

- ABN (non-profiled population) or TAU (profiled population) of business
- Type of Legal Organisation (TOLO)
- Standard Institutional Sector Classification Australia (SISCA)
- Employment size
- Industry (ANZSIC)
- Business turnover.

DATA CLEANING

Data cleaning was undertaken on the PIT data in order to remove duplicate records, remove invalid PAYG records (jobs with less than \$1 in gross payment), and derive data items which aligned with ABS standards and classifications. Duplicate records were identified as those where all variables were identical. Demographic variables (e.g. age and sex) were checked to ensure that they were referenced to 30 June 2012. Variables such as occupation were checked to ensure that they adhered to the ABS classifications and any erroneous or invalid codes were removed. After this cleaning, there were 12,734,746 records on the Client Register and Client Dataset (combined), and 13,316,438 records on the Individual PAYG Dataset.

Negligible data cleaning was required on the EABLD extract. After this cleaning, there were 6,917,943 records on the EABLD extract.

Integration Methodology

INTEGRATION METHODOLOGY

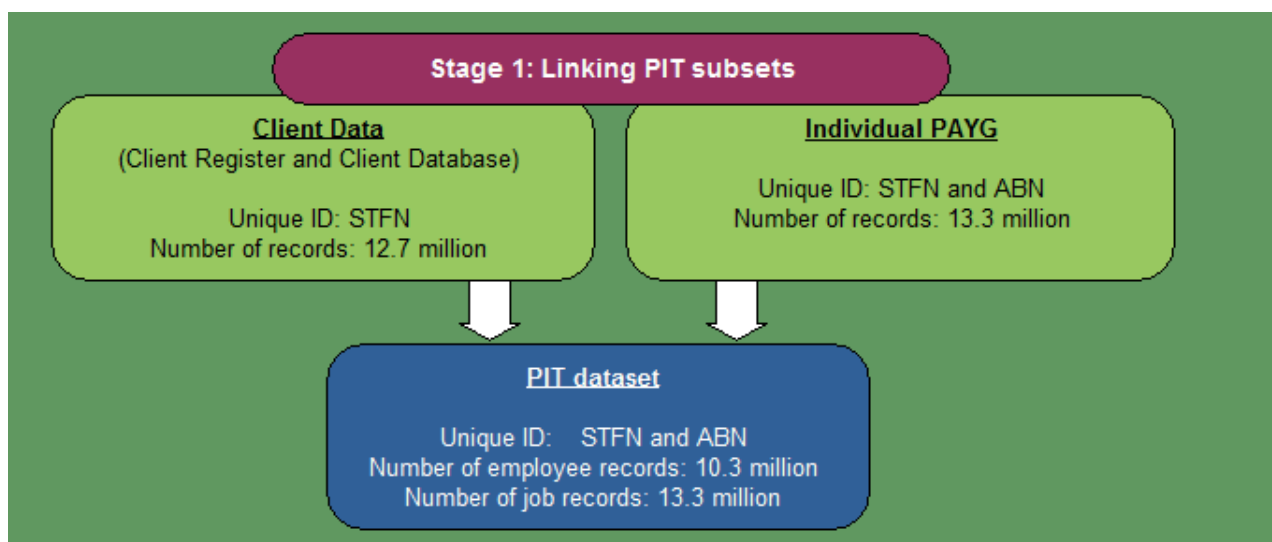
The LEED Foundation Projects Integrated Dataset was created through a two stage process. The first stage involved linking the three PIT subsets together, and the second stage involved integrating the PIT dataset with the EABLD extract.

STAGE 1: LINKING PIT DATASETS

The first stage of data integration involved linking the cleaned PIT subsets (Client Register, Client Dataset and Individual PAYG Dataset) together using STFAN as the linking key. Once these person level and job level data were linked, it was possible to identify persons who neither reported earnings on an ITR nor had an Individual PAYG summary during the 2011-12 financial year. These persons were deemed not to be employees in 2011-12 (out of scope) and were removed.

The resulting PIT dataset contains 10,334,718 employee records and 13,316,438 job records. Of all employee records, 94.4% have a corresponding job record. The 5.6% of employees who did not have a corresponding job record represent those employees who did not have an Individual PAYG summary for the reference period (such as persons who worked for cash in hand) but who still lodged an ITR and reported earnings. Of the job records, less than 0.001% did not have a corresponding employee record. As the linking variable is encrypted, it is impossible to determine whether any of the unlinked job records failed to link due to an error in the linking key, however the low number of unlinked jobs suggests that errors are minimal.

Diagram 2: Linking PIT datasets



STAGE 2: INTEGRATING PIT DATASET WITH EABLD

The second stage of data integration involved integrating the PIT dataset with the EABLD extract using ABN as the linking key. At this stage of the integration process, it was possible to identify businesses which did not have any PIT records linked to them. These businesses were deemed to be non-employing businesses in 2011-12 (out of scope) and were removed. As a result of this process, 683,331 business records were identified as employing businesses in the 2011-12 financial year.

Integrating the PIT dataset with the EABLD extract involved different processes for businesses in the non-profiled and profiled populations.

Non-profiled population

The PIT dataset was linked to non-profiled businesses on the EABLD extract using ABN as the linking key. This resulted in approximately 51% of jobs and 51% of employees being linked to a business in the non-profiled population.

Profiled population

Approximately 49% of jobs and 53% of employees were linked to a business in the profiled population. The EABLD extract contains information about profiled businesses at the TAU level. Because a profiled business may contain multiple TAUs, and because each ABN linked to this business can operate (either wholly or partially) in a selection of these TAUs, it was not possible to link ABNs of jobs directly to the TAUs of the businesses.

Rather than attempt to present the full complexity of businesses, it was decided to provide a single set of business information per job to allow for the effective calculation of employee and job statistics. In order to do this, the following *ABN to TAU mapping* process was employed to link each ABN to a single TAU within the EG. Linking entire ABNs to single TAUs allowed cohorts of employees working in the same business to remain linked together for microdata analysis. Other approaches such as linking individual jobs to TAUs may provide alternative means to perform this linking in future.

ABN to TAU mapping

For each ABN on the PIT dataset which linked to a business in the profiled population (i.e. an EG, represented by one or more TAUs on the EABLD extract), the ABN to TAU mapping process provided a link from an ABN to a single TAU. Each job associated with that ABN was linked to that

TAU. This process is based on information collected by the ABS as part of the ABSBR business profiling.

Of the 56,549 ABNs in the profiled population, 51% linked to a single TAU and further mapping was not required.

Of the remaining 49% of ABNs, approximately 0.5% could link to multiple TAUs, while the other 48.5% may link to one or multiple TAUs. Information from the ABS business profiling process in 2014 indicates that the overwhelming majority of ABNs (over 95%) in the profiled population map to a single TAU, but this information was not available for 2011-12.

A two-step process was developed to assign each remaining ABN to a single TAU. The first step was to calculate allocation weights (aw) between 0 and 99 for each potential combination of ABN and TAU. Each aw represents the probability that an ABN links to a particular TAU within the EG. The probabilities are based on the distribution of employees and income within the EG.

For example an ABN which may link to two TAUs (on a 70:30 probability ratio) would be represented by two ABN-TAU pairs:

- ABN:TAU 1, $aw = 70$
- ABN:TAU 2, $aw = 30$

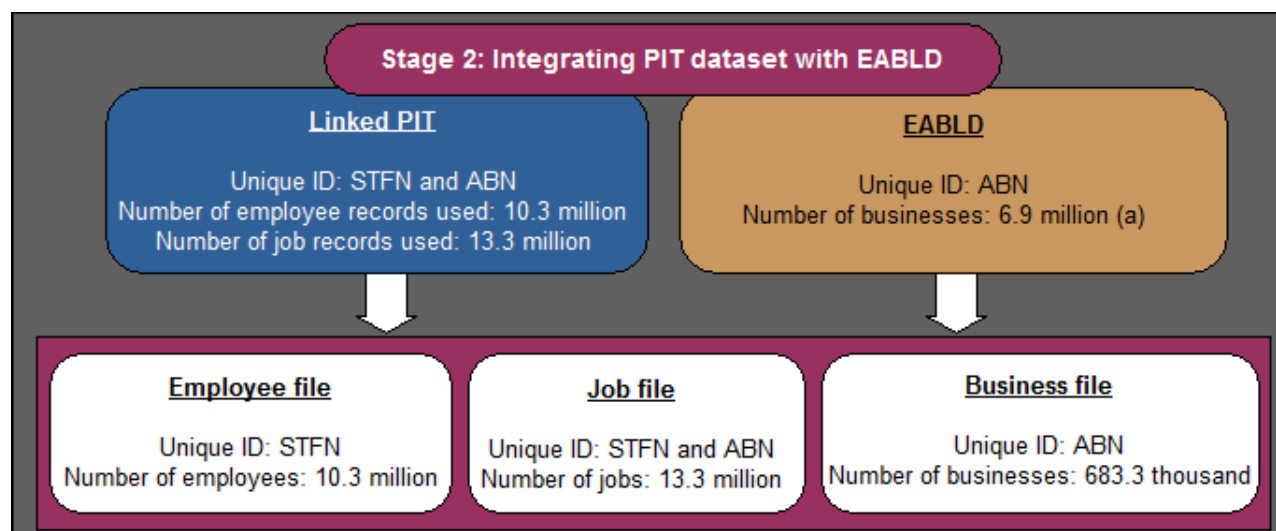
The second step was to assign each ABN to a single TAU in the EG. This was done by selecting, for each ABN, one of these ABN-TAU pairs. The likelihood of an ABN-TAU pair being selected was governed by an aw . ABN-TAU pairs with higher aw were more likely to be selected than those with lower aw . In the example above, the first ABN-TAU pair would have a 70% chance of being selected ($aw=70$), while the second pair would have a 30% chance ($aw=30$).

This approach assigned each ABN in the profiled population to a single TAU. As a result, each TAU may have 0, 1 or many ABNs linked to it. In addition, each EG can be represented by one or multiple TAUs, although not every TAU within each EG was necessarily selected (e.g. non-employing TAUs).

Although some ABNs and jobs (and therefore employees) may be allocated to TAUs in which they do not operate, this should have minimal impact on the experimental statistics at the aggregate level (for example the distribution of business-level variables such as Industry).

The ABN to TAU mapping process would benefit from improved coverage and quality of information collected as part of the ABSBR business profiling.

Diagram 3: The Integrated Dataset



(a) This includes all ABNs ever registered up to and including the 2011-12 financial year and does not reflect those active in the reference period.

INTEGRATED DATASET

The Integrated Dataset is comprised of three main subsets (files) representing three separate domains in the linked employer-employee data. These are employee-level data (the Employee File), job-level data (the Job File), and business-level data (the Business File). These files are linked together using unique keys where a link is possible, or left unlinked where no link could be made.

Employee File

The Employee File contains data relating to each employee. This includes demographic and aggregate earnings information, and selected information about jobs held. The Employee File contains all of the data items from the Client Register and Client Dataset extracts (see Data Sources), as well as the following derived data items:

- Earnings
- Items calculated using the Job File
 - Industry of main job
 - Number of jobs
 - Multiple job holder status
 - Number of concurrent jobs (for multiple job holders).

Job File

The Job File includes data relating to each job. This includes unique identifiers for employees and businesses, information about each job, whether the jobs is held concurrently with other jobs, and information about the business to which each job links. The Job File contains all of the data items on the Individual PAYG Dataset and the EABLD extract (see Data Sources), as well as the following derived data items:

- Occupation in main job (from the Employee File)
- Main job
- First job (for multiple job holders)
- Second job (for multiple job holders).

Business File

The Business File includes information relating to each business to which an Individual PAYG Dataset record is linked. The Business File contains all of the data items on the EABLD extract (see Data Sources).

INTEGRATION RESULTS

At the completion of the linking process:

- the Integrated Dataset contained 10,334,718 employees, 683,331 businesses, and 13,316,438 jobs;
- 9,751,414 employees (94%) from the Employee File were linked to a job on the Job File;
- 13,316,363 jobs (more than 99%) from the Job File were linked to an employee on the Employee File;
- 13,303,850 jobs (more than 99%) from the Job File were linked to a business on the Business File. Of these jobs:
 - 6,746,293 jobs (51%) were linked to a business in the non-profiled population;
 - 6,557,557 jobs (49%) were linked to a business in the profiled population;
- 675,571 businesses (99%) were in the non-profiled population, and were linked to 5,278,708 employees (51%) on the Employee File; and

- 7,760 businesses (1%) were in the profiled population, and were linked to 5,514,407 employees (53%) on the Employee File.

As some employees had more than one job during the reference period, these employees may link to more than one business on the Business File.

At the completion of the linking process, there were a number of unlinked records which are still included in the Integrated Dataset:

- 583,304 employees (6%) could not be linked to a job;
- 75 jobs (less than 0.001%) could not be linked to an employee; and
- 12,588 jobs (less than 0.1%) could not be linked to a business.

These unlinked records are due to:

- Employees reporting earnings on their ITR without a corresponding Individual PAYG summary (e.g. persons who worked for cash in hand); and
- Potential errors by employers on the Individual PAYG summary impacting the linking keys - Scrambled Tax File Number (STFN) and Australian Business Number (ABN).

Unlinked ABNs were examined and approximately 36% were found to be invalid. These errors are likely the result of erroneous data (e.g. typographical errors or Australian Company Numbers in place of ABNs) entered by employers on an Individual PAYG summary.

INTEGRATED DATASET COMPLETENESS

The following section highlights the completeness of some key data items on the Integrated Dataset.

Table 1: Inadequately defined, not stated or missing information

	Data items	Number of Records	
		('000)	%
Employee File	Geography – State and territory	189.3	1.8
	Geography – Statistical Area Level 4	211.7	2.1
	Occupation of main job	425.2	4.1
Job File	One date in PAYG	126.3	0.9
	Both dates in PAYG	98.2	0.7
Business File	Employment size	39.3	5.8
	Business turnover	17.7	2.6
	Industry	8.0	1.2
	Institutional sector	8.0	1.2

As seen in Table 1, there is a minimal amount of data that is unable to be defined or is missing following the minimal cleaning applied to the Integrated Dataset.

For further information on the distribution of key data items on the Integrated Dataset, see Appendix 1.

DATA CLEANING

Employee records with missing or extremely large, erroneous earnings (as reported on their ITR) impact on the detailed analysis of earnings. Limited cleaning was performed on the Integrated Dataset in order to adjust for missing earnings data, and to amend or remove extreme outliers.

Earnings cleaning

Employee records with \$0 earnings on an ITR (0.7%) were edited. The gross payments from each job (as recorded on Individual PAYG summaries) were aggregated, and this value replaced the \$0 earnings value. As earnings are comprised of amounts other than gross payments, these records likely have slightly understated earnings compared to what would otherwise be reported on an ITR.

Outlier cleaning

Extreme earnings outliers were identified within each employee age and occupation strata. These employee records were further examined by confronting their earnings data against the aggregate of all gross payments for that employee. Records which either had no gross payment data, or for whom the differences were not reconcilable, were identified as extreme outliers and amended in one of two ways. Firstly, if gross payment data existed, the earnings value was deleted and replaced with the aggregate gross payments amount for that employee. Secondly, if gross payment data was not available, the earnings value was excluded from the calculation of the experimental statistics on employee earnings (although the employee record still contributed to statistics on employees). This cleaning affected in total approximately 0.001% of employees.

Summary of Results

SUMMARY OF RESULTS

OVERVIEW

The LEED Foundation Projects have produced experimental statistics on employees and earnings for the 2011-12 financial year. The key findings are:

- there were 10.3 million employees;
- the median and mean earnings for all employees were \$45,869 and \$55,678 respectively;
- there were 5.4 million male employees (52%) and 5.0 million female employees (48%);
- the median and mean earnings for males were \$55,470 and \$66,994 respectively, while for females they were \$37,726 and \$43,489.

The LEED Foundation Projects experimental statistics provide new information on jobs and multiple job holders for the 2011-12 financial year which is not currently available from ABS statistics. The key findings are:

- there were 13.3 million jobs;
- the median and mean gross payments received for all jobs were \$26,134 and \$37,961 respectively;
- there were 6.8 million jobs held by males (51%) and 6.5 million by females (49%);
- the median and mean gross payments for males were \$34,495 and \$46,386 respectively, while for females they were \$20,429 and \$29,161;
- there were 1.9 million multiple job holders (individuals with two or more concurrent jobs) whose median and mean earnings were \$38,892 and \$49,875 respectively, and of which 53% were female and 47% were male.

The LEED Foundation Projects experimental statistics are based on administrative data and may include non-sampling error. There are a number of data considerations that users should be aware of when interpreting or analysing the experimental statistics (see Explanatory Notes, paragraphs 31-69).

DISTRIBUTION OF EMPLOYEES & EARNINGS

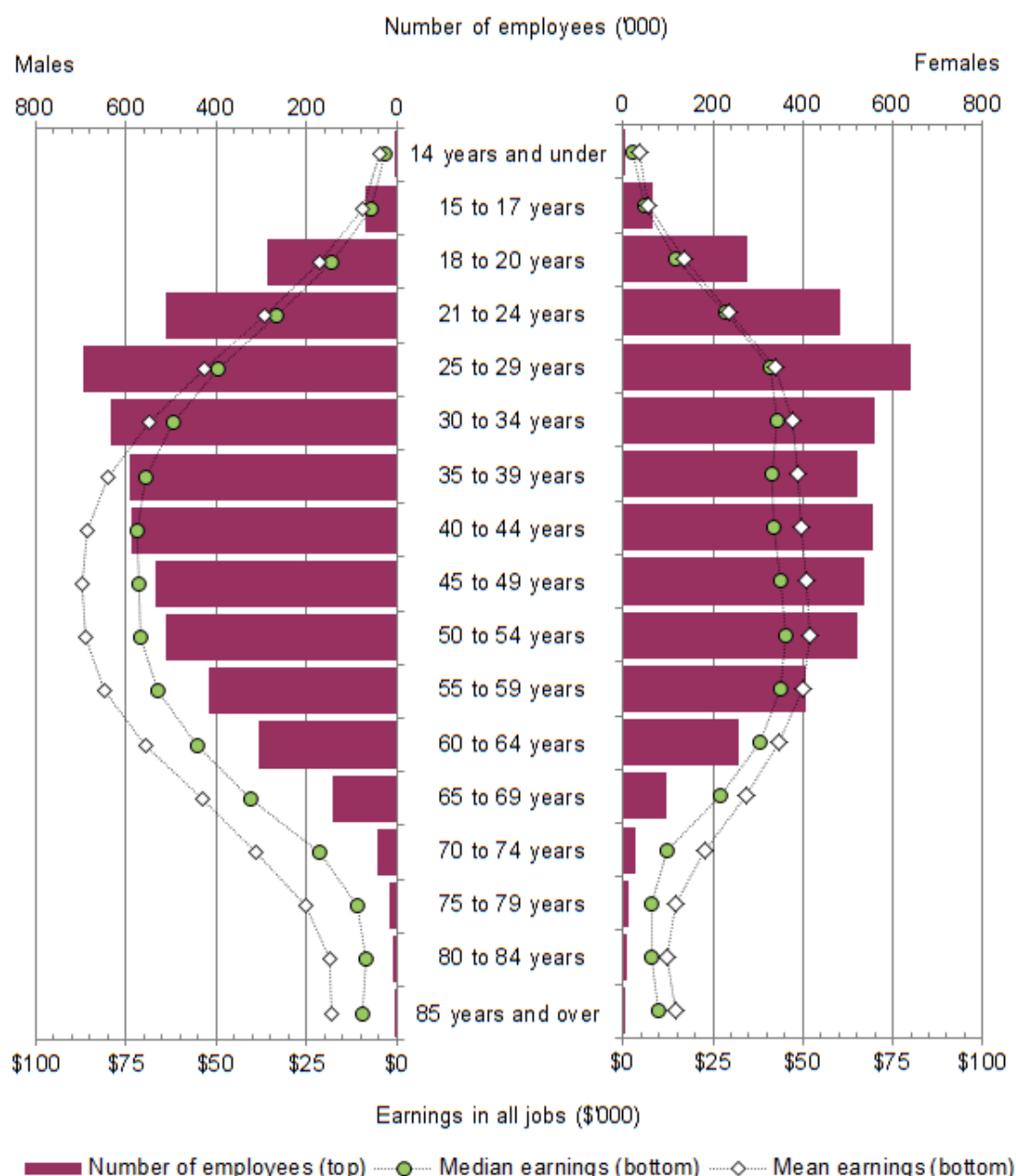
In the 2011-12 financial year there were 5.4 million male employees with median and mean earnings of \$55,470 and \$66,994 respectively. There were a total of 5.0 million female employees with median and mean earnings of \$37,726 and \$43,489 respectively.

Age and sex

For both male and female employees, the median earnings in all age groups are less than the mean earnings, reflecting positively skewed earnings distributions. Employees aged between 25 and 29 years were the largest group (1.3 million) and had the lowest difference (6.6%) between their median and mean earnings (\$45,215 and \$48,193 respectively). This indicates that employee earnings in this age group were less skewed compared to other age groups in the distribution. In most age groups, earnings were relatively more skewed for male employees than for female employees, however in the 35 to 39 years age group, earnings were more skewed for female employees (17% difference) than for male employees (15% difference).

Median and mean earnings for the 25 to 29 years age group differ from the median and mean earnings for the 45 to 49 years age group. For males, they differ by \$22,053 and \$33,535 respectively, while for females they differ by \$2,817 and \$8,602 respectively. For male employees, median earnings begin to decrease in the 45 to 59 years age group, while mean earnings begin to decrease in the 50 to 54 years age group. For female employees, median earnings begin to decrease in both the 35 to 39 and the 55 to 59 years age groups and mean earnings begin to decrease in the 55 to 59 age group. For further information, see the Aggregate Experimental Statistics Data Cube, Table 1 in the Downloads tab.

Graph 1: Number of employees and earnings in all jobs, by age group and sex, 2011-12



State and territory

In the 2011-12 financial year, New South Wales had the highest number of employees (30.8%), followed by Victoria (24.5%) and Queensland (20.0%). The Australian Capital Territory and Western Australia had the highest mean earnings (\$64,685 and \$62,732 respectively). However, median earnings for Western Australia (\$50,285) were 25% lower than the mean, while in the Australian Capital Territory median earnings (\$59,523) were 9% lower, reflecting different skewness in the earnings distribution. For further information, see the Aggregate Experimental Statistics Data Cube, Table 1 in the Downloads tab.

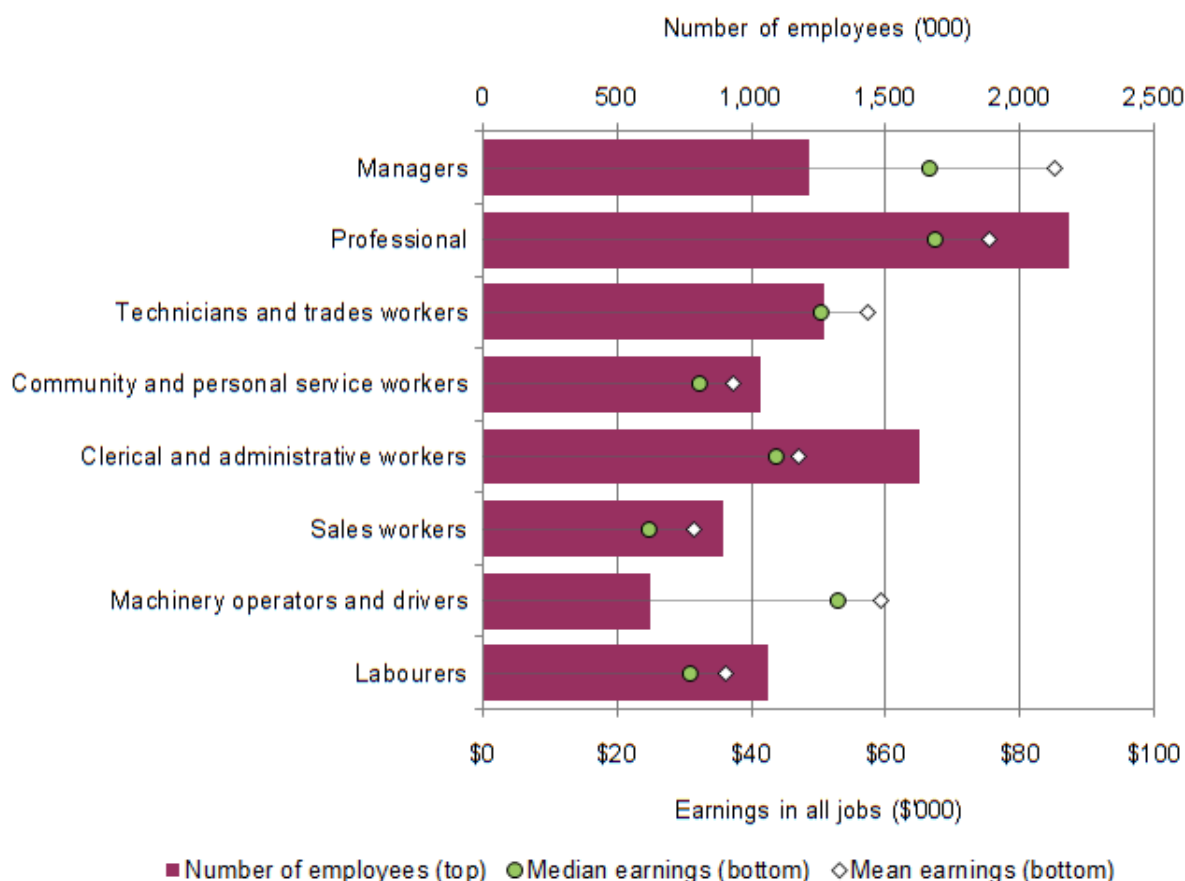
Graph 2: Number of employees and earnings in all jobs, by state and territory, 2011-12



Occupation

Among all major occupation categories, Professionals had the highest median earnings (\$67,512) followed by Managers (\$66,652). However, Managers' mean earnings (\$85,286) were 28% higher than their median, while Professionals' mean earnings (\$75,541) were 12% higher. This reflects a more skewed earnings distribution for Managers than Professionals. For further information, see the Aggregate Experimental Statistics Data Cube, Table 1 in the Downloads tab.

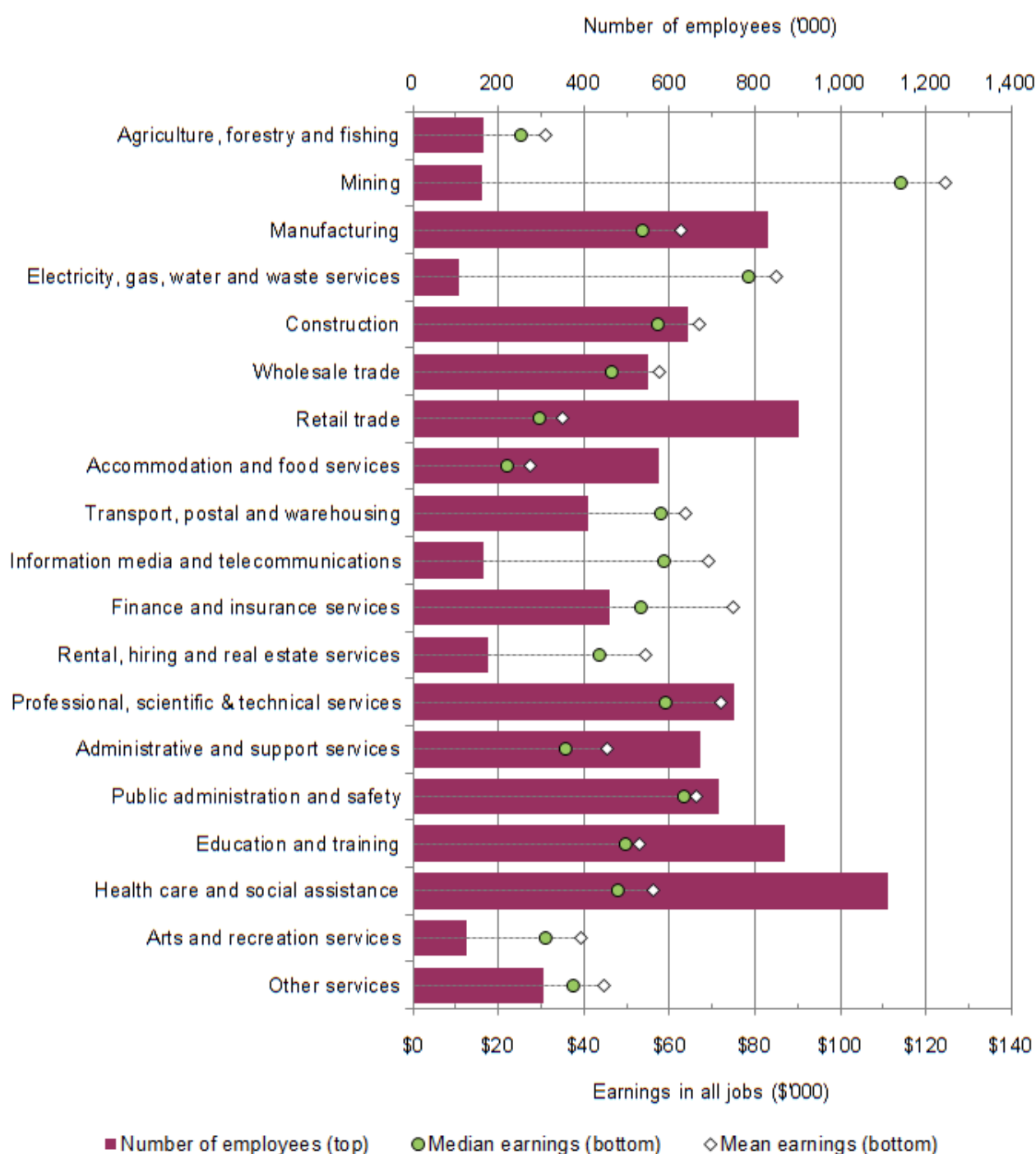
Graph 3: Number of employees and earnings in all jobs, by occupation in main job, 2011-12



Industry

Health care and social assistance had the highest number of employees (11%) followed by Retail trade (8.7%) and Education and training (8.4%). Employees in Mining had the highest median and mean earnings (\$114,053 and \$124,589 respectively), followed by Electricity, gas, water and waste services (median \$78,730 and mean \$85,235). On the other hand, Accommodation and food services had the lowest median and mean earnings (\$22,012 and \$27,549 respectively). Employees in Finance and insurance services had the most skewed earnings data, with a 40% difference between median and mean earnings (\$53,530 and \$74,892 respectively). In contrast, employees in Public administration and safety had the least skewed earnings data, with a 4% difference between their median and mean earnings (\$63,632 and \$66,317 respectively). For further information, see the Aggregate Experimental Statistics Data Cube, Table 1 in the Downloads tab.

Graph 4: Number of employees and earnings in all jobs, by industry of main job, 2011-12



DISTRIBUTION OF JOBS & GROSS PAYMENTS

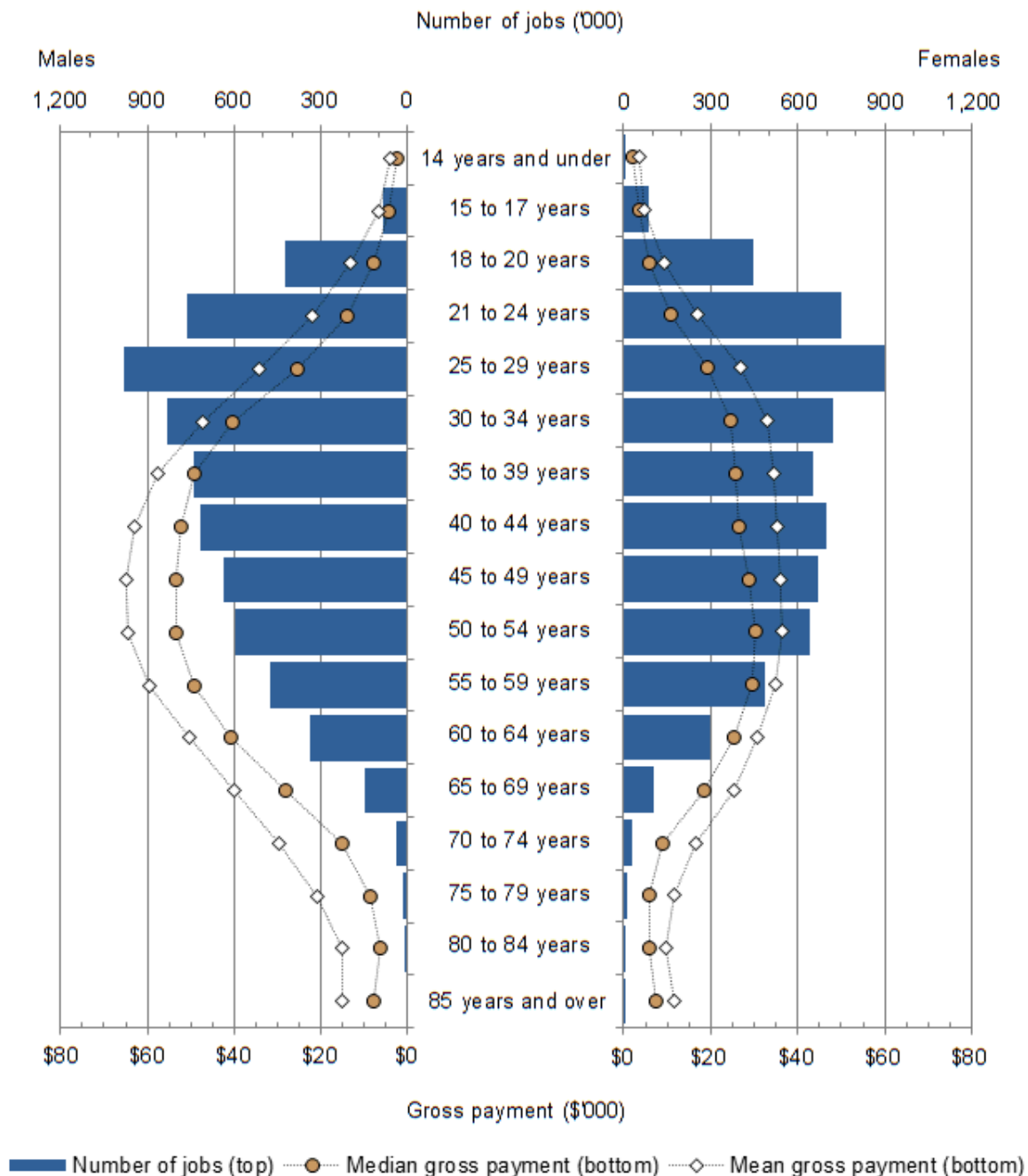
As a result of the LEED Foundation Projects, the ABS is able to produce experimental statistics on filled jobs for the first time.

In the 2011-12 financial year there were 13.3 million jobs in Australia. The median gross payment for all jobs was \$26,134 and the mean was \$37,961. Approximately 51% of all jobs were occupied by males, with median and mean gross payments of \$34,495 and \$46,386 respectively. Approximately 49% were occupied by females, with median and mean gross payments of \$20,429 and \$29,161 respectively.

Age and sex

Male employees had higher gross payments in all jobs across all age groups. Their highest median and mean gross payments were in the 45 to 49 years age group (\$53,497 and \$64,904 respectively), whereas for females they were in the 50 and 54 years age group (\$30,436 and \$36,357 respectively). The gross payments were relatively more skewed for female employees aged between 25 and 44 years than for male employees. For further information, see the Aggregate Experimental Statistics Data Cube, Table 2 in the Downloads tab.

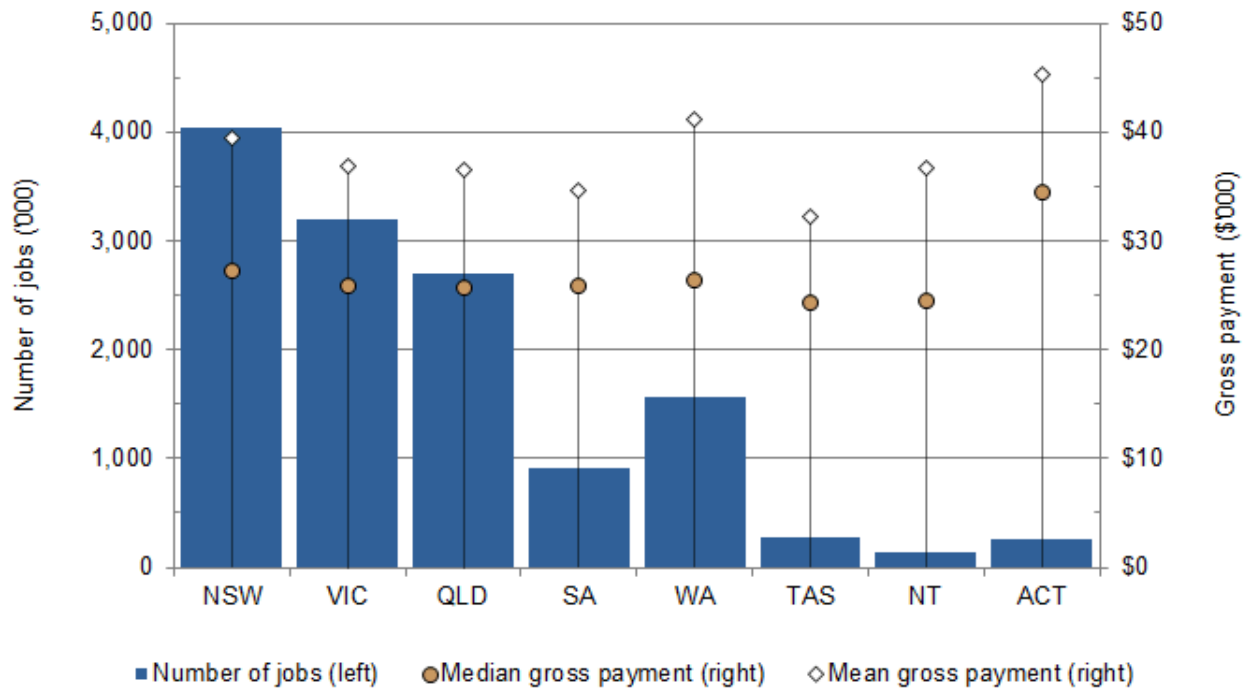
Graph 5: Number of jobs and gross payment, by age group and sex, 2011-12



State and territory

There were 4.0 million jobs in New South Wales (30%), followed by 3.2 million in Victoria (24%) and 2.7 million in Queensland (20%). As for earnings, the highest mean gross payments were in the Australian Capital Territory (\$45,276) followed by Western Australia (\$41,138). The median gross payment for Western Australia (\$26,354) was 36% lower than the mean, while in the Australian Capital Territory the median gross payment (\$34,520) was 24% lower, reflecting different skewness in the gross payments distribution. For further information, see the Aggregate Experimental Statistics Data Cube, Table 2 in the Downloads tab.

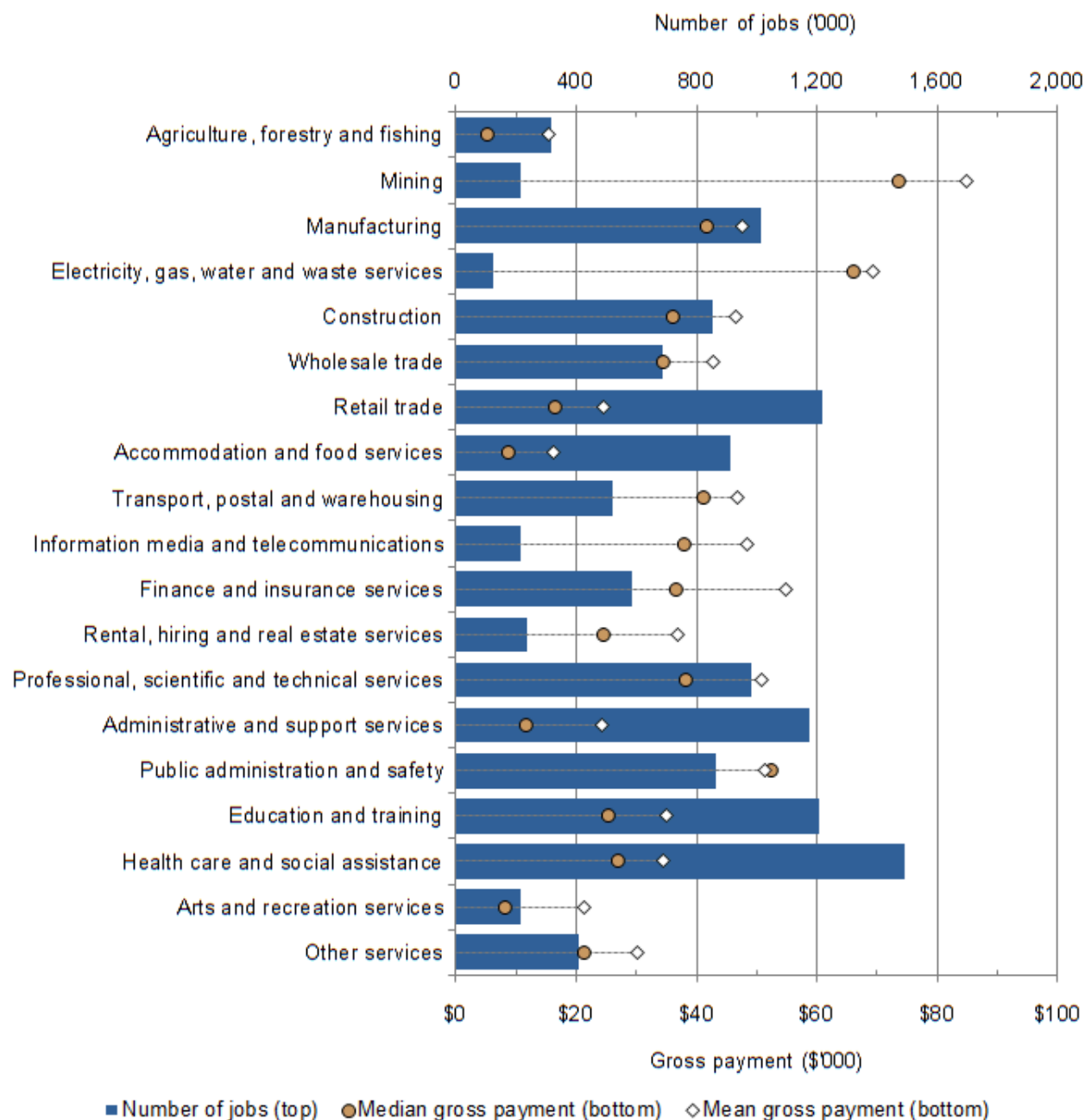
Graph 6: Number of jobs and gross payment, by state and territory, 2011-12



Industry

Health care and social assistance had the highest number of jobs (11%) followed by Retail trade (9.2%) and Education and training (9.1%), mirroring the number of employees (see Graph 4). Jobs in the Mining industry had the highest median gross payments (\$73,645) followed by those in Electricity, gas, water and waste services (\$66,158). For further information, see the Aggregate Experimental Statistics Data Cube, Table 2 in the Downloads tab.

Graph 7: Number of jobs and gross payment, by industry, 2011-12



Sector and employment size

The Private sector had 78% of all jobs, with the remaining 22% in the Public sector. Public sector employees had higher gross payments. In the 2011-12 financial year, approximately 50% of jobs were in businesses with 200 or more employees, and employees in these large businesses had the highest median and mean gross payments. For further information, see the Aggregate Experimental Statistics Data Cube, Table 2 in the Downloads tab.

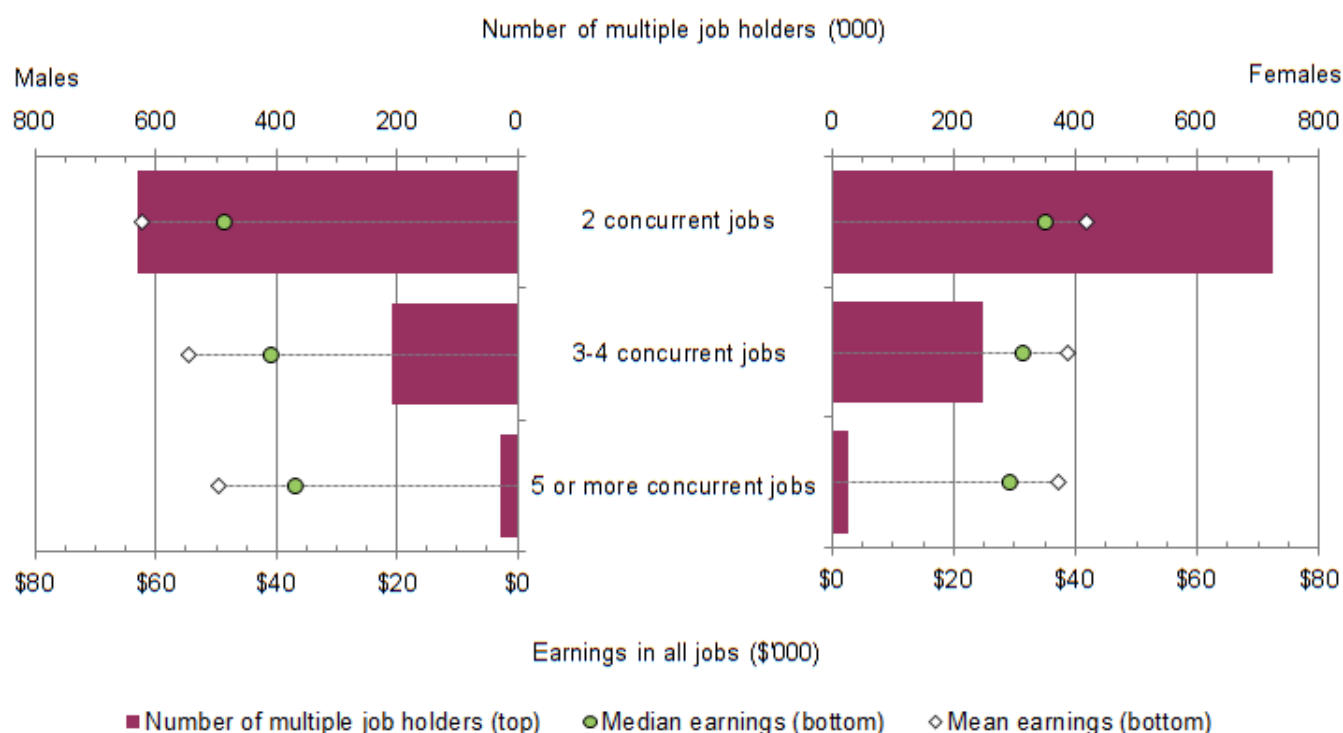
Table 2: Number of jobs and gross payment, by type of legal organisation and employment size, 2011-12

	Number of jobs ('000)	Median gross payment \$	Mean gross payment \$
Type of legal organisation			
Private sector entities	10 434.9	23 309	36 290
Public sector entities	2 865.0	38 971	44 069
Employment size			
Fewer than 5 employees	1 423.0	16 640	27 727
5-19 employees	2 008.8	18 279	28 701
20-199 employees	3 276.9	21 383	33 588
200 or more employees	6 595.1	35 066	45 192

Multiple job holders

There were 1.9 million employees who had multiple (concurrent) jobs at any point during the 2011-12 financial year. Of these, 53% were female and 47% were male. Approximately 73% of the multiple job holders had two concurrent jobs, while 24% had three to four concurrent jobs during the financial year. For further information, see the Aggregate Experimental Statistics Data Cube, Table 3 in the Downloads tab.

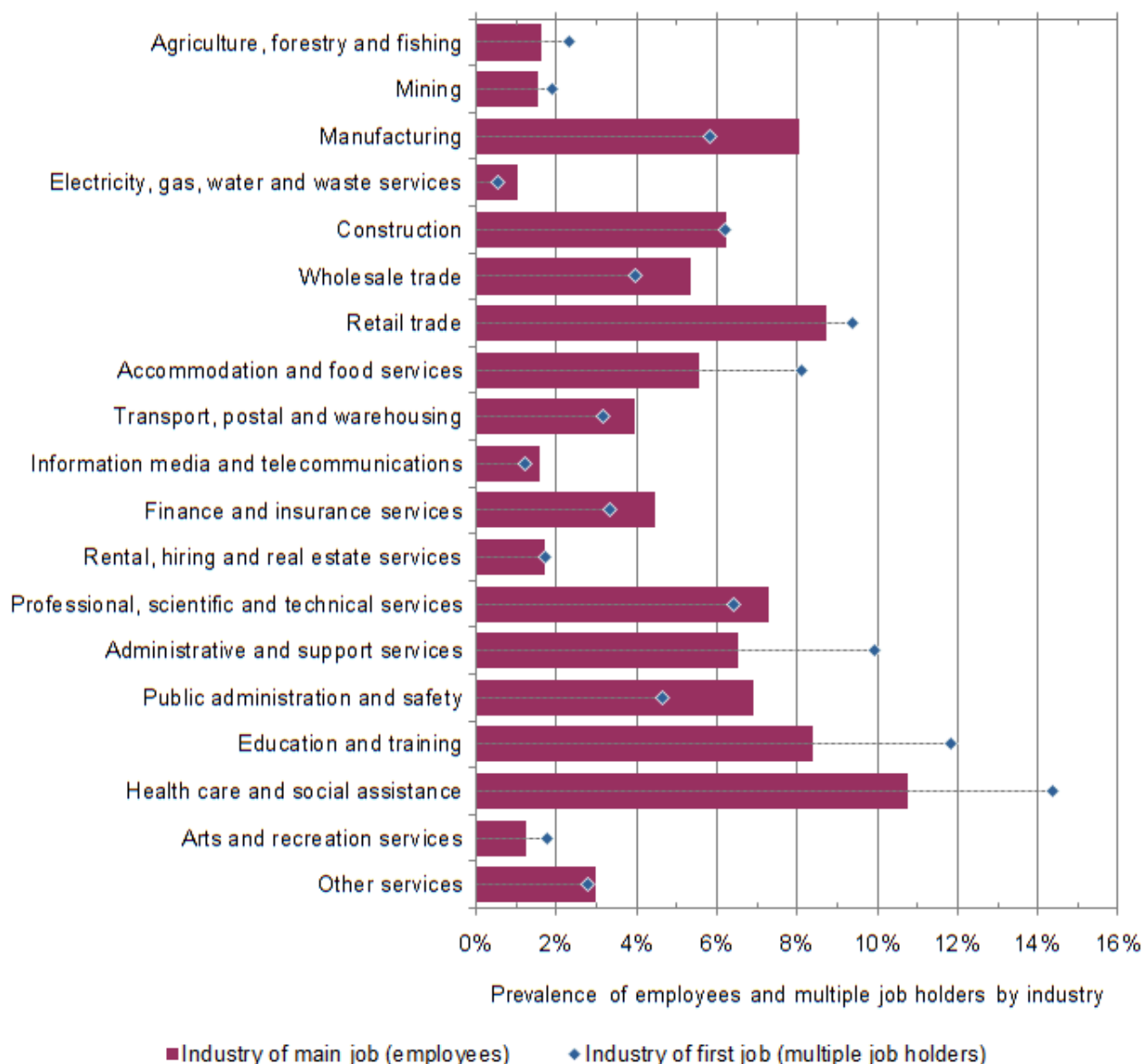
Graph 8: Number of multiple job holders and earnings in all jobs, by sex, 2011-12



The graph below compares the distribution of industry of first job for multiple job holders (the job with highest gross payment among concurrent jobs) with industry of main job for all employees (the job with highest gross payment).

The distribution of the industry of first job for multiple job holders differs from the industry of main job for all employees. Health care and social assistance, Education and training, and Administrative and support services were the industries in which multiple job holders were most concentrated (14%, 12% and 10% respectively). However, employees were less concentrated in these same industries (11%, 8.4% and 6.5% respectively). For further information, see the Aggregate Experimental Statistics Data Cube, Table 4 in the Downloads tab.

Graph 9: Prevalence of multiple job holding by industry, 2011-12



LIMITATIONS OF THE EXPERIMENTAL STATISTICS

The construction of the experimental statistics on employee earnings and jobs identified a number of data considerations which need to be taken into account when interpreting these statistics. These include using information as defined by and reported to the ATO; earnings being comprised only of reported amounts; multiple job holder status being determined on reported dates information; and the allocation of industry for jobs in profiled businesses being contingent on the ABN to TAU mapping.

Care should be taken when interpreting or analysing the experimental statistics. For further information on data considerations see Explanatory Notes (paragraphs 31-69).

COHERENCE OF EXPERIMENTAL STATISTICS WITH ABS SURVEY COLLECTIONS

Overall, the experimental statistics were found to be broadly coherent with current ABS household and business survey estimates. However, some differences were identified between the experimental statistics and survey estimates due to differences in scope, sample design, collection methodology and processing approaches. Moreover, the Integrated Dataset used to construct the experimental statistics is based on data collected for administrative purposes, whereas ABS collections are explicitly designed to create statistical outputs.

For further information on the coherence of the experimental statistics with ABS estimates see

Future Directions

FUTURE DIRECTIONS

ACCESS TO MICRODATA

The ABS will provide access to a sample of the Integrated Dataset through *Microdata: Employee Earnings and Jobs, Australia, 2011-12* (cat. no. 6311.0.55.001). The microdata will be made available as a de-identified unit record file, released with the approval of the Australian Statistician, via the ABS Data Laboratory. To mitigate risks of disclosure (of persons and businesses) the microdata file will be a sample of the Integrated Dataset, representative of the in-scope employee level records. The size of the sample is expected to be approximately 10% (about 1 million employees). Further information regarding the microdata product will be released on the ABS website in due course.

TRANSFORMING FOR THE FUTURE

Over the next five years, the ABS will fundamentally transform across all aspects of the organisation to achieve the vision of unleashing the power of statistics for a better Australia. A key focus during the transformation is to better utilise data collected for administrative purposes, and improving the availability and use of Australia's statistical assets. A future LEED is one such statistical solution to meet increasing demands to deliver more and improved information on the Australian labour market.

As outlined in the Introduction, a future LEED will provide a longitudinal view of linked employer and employee data, based on person and business level data from the ATO. The creation of an enduring LEED will aid in building the capacity to undertake analysis of the drivers of labour market change, firm-level productivity, as well as sustainable regional economies. It will complement the current labour statistics produced by the ABS, and improve the evidence base for policy development and evaluation, leading to more targeted expenditure of government funds.

The ABS is exploring the viability of producing an ongoing tax-based LEED as part of the transformation agenda. While a future LEED is seen as a valuable statistical asset, the ABS requires users to clearly articulate how this information will provide new insights on key public policy issues. As part of this process, the ABS will be engaging with key stakeholders to gather user requirements so that a fit for purpose new solutions project is developed.

REFINED INTEGRATION METHODOLOGY

The construction of the Integrated Dataset and the experimental statistics on employee earnings and jobs required a number of assumptions to be made. These include assumptions about the person level derived data items (such as Main job and Earnings), and the process for allocating each ABN within a complex business to a single TAU within the EG. Further investigations into the assumptions and apportioning methods should be pursued as part of the future LEED development. This could be achieved through gaining a better understanding of the relationships in the PIT data and related ABS data sources; securing additional data items in the PIT data (such as Occupation labels); as well as improving the coverage and quality of the ABN-TAU mapping through the ABSBR profiling process. In addition, the ABS could partner with key stakeholders to gather expertise on data cleaning and editing strategies, as well as the integration methodology to enhance the utility of the LEED. Further to this, the ABS has already committed to refine the integration methodology used to construct the EABLD, which would also support the development of the future LEED.

INVITATION TO COMMENT

If you wish to provide feedback on the LEED Foundation Projects and the creation of an enduring LEED as part of the ABS' transformation agenda, please feel free to contact the Labour and Income Branch on (02) 6252 7206 or via labour.statistics@abs.gov.au. The ABS Privacy Policy outlines how the ABS will handle any personal information that you provide to us.

About this Release

This information paper summarises the process used to construct experimental statistics on employee earnings and jobs from administrative data for the 2011-12 financial year. The experimental statistics are based on linked employer-employee data constructed by integrating Personal Income Tax data sourced from the Australian Taxation Office (ATO) with business level data from the ABS Expanded Analytical Business Longitudinal Database (EABLD). The EABLD integrates business level data collected by the ABS and administrative data sourced from the ATO. Information on data sources are available in this paper. A data cube of the aggregate experimental statistics on employee earnings and jobs is available in this release.

Explanatory Notes

Explanatory Notes

EXPLANATORY NOTES

INTRODUCTION

1 The Australian Bureau of Statistics (ABS) has constructed new experimental statistics on employees, earnings, and jobs in the Australian labour market for the 2011-12 financial year. This release presents these new experimental statistics based on integrated person and business level taxation data from the Australian Taxation Office (ATO).

2 The experimental statistics were compiled from the Integrated Dataset produced by the LEED Foundation Projects. The Integrated Dataset was created by integrating:

- person level data from the ATO Personal Income Tax (PIT) dataset; and
- business level taxation data from the ABS Expanded Analytical Business Longitudinal Database (EABLD). The EABLD is based on the ABS Business Register (ABSBR), and incorporates business taxation data from the ATO.

3 In undertaking these projects, the ABS complied fully with the High Level Principles for Data Integration Involving Commonwealth Data for Statistical and Research Purposes.

4 For further information on the LEED Foundation Projects see the project entry on the Public Register of Data Integration Projects on the National Statistical Services (NSS) website.

DATA SOURCES

Personal Income Tax data

5 The PIT data is sourced from the ATO.

6 The results of the LEED Foundation Projects are based, in part, on tax data supplied by the ATO to the ABS under the **Taxation Administration Act 1953**, which requires that such data is only used for the purpose of administering the **Census and Statistics Act 1905**. Any discussion of data limitations or weaknesses is in the context of using the data for statistical purposes, and is not related to the ability of the data to support the ATO's core operational requirements.

7 Legislative requirements to ensure privacy and secrecy of this data have been adhered to. In accordance with the **Census and Statistics Act 1905**, results have been confidentialised to ensure that they are not likely to enable identification of a particular person or organisation.

8 The data has been collected in compliance with Australian taxation laws. The unit record data was provided to the ABS for a variety of statistical purposes and so was not tailored specifically to this project. The unit record PIT dataset contains a range of key data items such as income and demographic data items such as age, sex and birth year. Information on the statistics contained in the dataset is generally available through the ATO website.

9 Data provided to the ABS by the ATO are from ITR processed up to 16 months after the end of the financial year (i.e. returns processed up to 31 October 2013 for the financial year ending 30 June 2012). Due to the identifying nature of the data it contains, access to all ATO datasets is strictly regulated by the ATO. Both the ATO and the ABS handle personal information in accordance with the Australian Privacy Principles contained in the **Privacy Act 1988**.

10 The LEED Foundation Projects used an extract of the PIT data for the 2011-12 financial year containing selected variables (see Data Sources).

11 According to taxation laws, individuals whose income is below a certain threshold are in many instances not required to submit tax returns. However, amendments to the taxation laws can significantly alter the information that is required to be reported in the ITRs. Statistics derived from the PIT dataset will be influenced by tax regulation changes. The tax-free threshold in the 2011-12 financial year was \$6,000.

Expanded Analytical Business Longitudinal Database

12 In partnership with the Department of Industry, Innovation and Science, the ABS is developing an enduring firm level statistical asset that will increase the capacity of the research community to undertake firm level analysis of micro-economic drivers of performance, competitiveness, and productivity, and improve the evidence base for policy development and evaluation, leading to more targeted expenditure of government funds.

13 The EABLD integrates administrative data from the ATO with collected ABS survey data for all economically active businesses in the Australian economy from 2001-02 to 2012-13.

14 The LEED Foundation Projects used an extract of the EABLD for the 2011-12 financial year containing selected variables. The EABLD extract contains a record for every business ever registered up to and including the 2011-12 financial year.

15 The EABLD is comprised of two business populations, as described below:

- *Non-profiled population (simple businesses)* - The majority of businesses in the EABLD have simple structures and the unit registered for an ABN will satisfy ABS statistical reporting requirements. These businesses form the non-profiled population. The ABN is the statistical unit used in the LEED Foundation Projects to represent businesses in the non-profiled population.
- *Profiled population (complex businesses)* - For those businesses where the ABN is not considered suitable for ABS statistical requirements, the ABS maintains its own units structure (the ABS Units Model) through direct contact with businesses. This population, known as the profiled population, consists typically of large, diverse and

complex structured businesses. For businesses in the profiled population, statistical units include the Enterprise Group (EG) and the Type of Activity Unit (TAU). The range of activities carried out across the EG can be very diverse. The TAU is established to represent a grouping of one or more businesses within the EG that cover all the operations within an industry sub-division and for which a basic set of financial, production and employment data can be reported. The TAU is the statistical unit used in the LEED Foundation Projects to represent businesses in the profiled population, such that each TAU of a complex business (EG) is considered to be a separate business.

16 For further information on the construction of the EABLD, refer to Information Paper: Construction of the Expanded Analytical Business Longitudinal Database, 2001-02 to 2012-13 (cat. no. 8171.0).

17 For further information on the ABS Units Model, refer to the Appendix of the Standard Economic Sector Classifications of Australia, 2008 (cat. no. 1218.0).

18 The results of the LEED Foundation Projects are based, in part, on Australian Business Register (ABR) data supplied by the Registrar to the ABS under **A New Tax System (Australian Business Number) Act 1999** which requires that such data is only used for the purpose of carrying out functions of the ABS. Any discussion of data limitations or weaknesses is in the context of using the data for statistical purposes, and is not related to the ability of the data to support the ABR's core operational requirements.

SCOPE

19 The LEED Foundation Projects capture information on all employee earnings and jobs in Australia throughout the reference period of 1 July 2011 to 30 June 2012.

20 The scope of the LEED Foundation Projects includes:

- all persons who were an employee at any point in the reference period as recorded on either an Individual Tax Return (ITR) or an Individual Pay As You Go (PAYG) summary.
- all jobs as reported in an Individual PAYG summary during the reference period; and
- all businesses which provided an Individual PAYG summary to an employee in the reference period.

COVERAGE

21 Coverage restrictions apply to the LEED Foundation Projects Integrated Dataset.

22 The LEED Foundation Projects' use of unique identifiers ensures that each individual is unlikely to be included more than once in the experimental statistics.

23 Employees who meet one of the following conditions will be partially excluded from the LEED Foundation Projects. For these employees, missing information from one source (e.g. missing PAYG data) will result in exclusion from certain statistics (e.g. Mean Gross Payments, or Number of Jobs).

1. Employees who did not report earnings on an ITR for any of the following reasons:
 - Did not submit an ITR for any of the reasons outlined on pages 6 and 7 of the Individual tax return instructions 2012 on the ATO website;
 - Did not submit an ITR for any other reason; or
 - Submitted an ITR but did not report their applicable earnings.
2. Employees who did not receive an Individual PAYG summary from an employer for any of the following reasons:
 - They worked for cash in hand or other payments not recorded on an Individual

PAYG summary;

- They conducted illicit activities not recorded on Individual PAYG summaries;
- They did not supply their Tax File Number (TFN) to their employer; or
- Any other reason.

24 No employing businesses were excluded on the basis of coverage.

INTEGRATION RESULTS

25 The **Integrated Dataset** is comprised of three main subsets representing three separate domains in the linked employer-employee data:

- Employee File - contains data relating to each employee from the PIT Client Register and Client Dataset;
- Job File - contains data relating to each job from the PIT Individual PAYG Dataset; and
- Business File - contains data relating to each business from the EABLD. Note, the Business File only contains information on businesses which could be linked to a job on the Job File.

26 These files are linked together using unique keys where a link is possible, or left unlinked where no link could be made. The linking keys are the Scrambled Tax File Number (STFN) and Australian Business Number (ABN).

27 At the completion of the linking process:

- the Integrated Dataset contained 10,334,718 employees, 683,331 businesses, and 13,316,438 jobs;
- 9,751,414 employees (94%) from the Employee File were linked to a job on the Job File;
- 13,316,363 jobs (more than 99%) from the Job File were linked to an employee on the Employee File;
- 13,303,850 jobs (more than 99%) from the Job File were linked to a business on the Business File. Of these jobs:
 - 6,746,293 jobs (51%) were linked to a business in the non-profiled population;
 - 6,557,557 jobs (49%) were linked to a business in the profiled population.
- 675,571 businesses (99%) were in the non-profiled population, and were linked to 5,278,708 employees (51%) on the Employee File; and
- 7,760 businesses (1%) were in the profiled population, and were linked to 5,514,407 employees (53%) on the Employee File. As some employees had more than one job during the reference period, these employees may link to more than one business on the Business File.

28 At the completion of the linking process, there were a number of unlinked records which are still included in the Integrated Dataset:

- 583,304 employees (6%) could not be linked to a job;
- 75 jobs (less than 0.001%) could not be linked to an employee; and
- 12,588 jobs (less than 0.1%) could not be linked to a business,

29 These unlinked records are due to:

- Employees reporting earnings on their ITR without a corresponding Individual PAYG summary (e.g. persons who worked for cash in hand); and
- Potential errors by employers on the Individual PAYG summary impacting the linking keys - STFN and ABN.

30 Unlinked ABNs were examined and approximately 36% were found to be invalid. These errors are likely the result of erroneous data (e.g. typographical errors or Australian Company Numbers in place of ABNs) entered by employers on an Individual PAYG summary.

DATA CONSIDERATIONS

31 There are a number of data considerations that users should be aware of when interpreting or analysing the experimental statistics.

Employees

32 Employees are defined as persons who worked for a private or public sector employer and received pay for the reference period in the form of wages or salaries, a commission while also receiving a retainer, tips, piece rates or payments in kind. Persons who operated their own incorporated enterprises with or without hiring employees are also included as employees.

33 The definition of employees used in the experimental statistics includes Owner Managers of Incorporated Enterprises (OMIEs) as they cannot be separately identified from other employees based on the PIT data. This aligns with the previous ABS Status in Employment classification. The current ABS Status in Employment classification separately identifies employees and OMIEs. The current classification was released in 2014. As such, coherence between ABS estimates of employees in 2011-12 and the experimental statistics is not affected by the inclusion of OMIEs as employees.

34 For further information on the ABS Status in Employment classification, see Standards for Labour Force Statistics, Issue for Dec 2014 (cat. no. 1288.0).

35 For the purpose of the LEED Foundation Projects, employees only contribute to the experimental statistics if they have a record on the Employee File. The small number of employees who have valid jobs on the Job File (and therefore are conceptually employees) but do not have a corresponding record on the Employee File, are excluded from the experimental statistics on employees (although they are included in the jobs statistics).

Earnings

36 Earnings are the gross amounts paid to employees for work done or time worked (including paid leave). Earnings is the aggregate of total payments (in cash and in kind) received by each employee in all of their jobs, as reported on an ITR.

37 The variables used in constructing earnings are all taken from the ITR and their total values summed together. They are as follows:

- Salary or wages
- Allowances, earnings, tips, directors fees etc.
- Employer lump sum payments
- Attributed personal services income
- Employee share schemes, total assessable discount amount
- Total reportable fringe benefits amounts
- Reportable employer superannuation contributions.

38 Irregular payments (e.g. bonuses, share scheme discounts, employer lump sum payments etc.) are part of the notional concept of earnings but for many ABS collections they are excluded to minimise seasonal variation. Because the reference period for the LEED Foundation Projects is the entire 2011-12 financial year rather than a point in time snapshot, any irregular payments made within the reference period are included.

39 For information on the ABS concepts and definitions related to earnings, see Chapter 12:

Employee Remuneration, Labour Statistics: Concepts, Sources and Methods, 2013 (cat. no. 6102.0.55.001).

40 For information on the concepts of earnings used in the ABS business and household surveys, see the feature article “Understanding Earnings in Australia Using ABS Statistics”, July 2014 issue, Australian Labour Market Statistics (cat. no. 6105.0).

41 It is important to note that earnings information included in the experimental statistics is for **all jobs** as reported to the ATO through an ITR. It is not possible to disaggregate all the components of earnings per job using the PIT data. The gross payment information is the **only** job level payment information which can be separately identified based on the PIT data (see Explanatory Notes, paragraph 61).

42 Cash in hand work or earnings within the tax free threshold (less than \$6,000) will not be included if no earnings were reported to the ATO through an ITR.

43 For the purpose of the LEED Foundation Projects, employees only contribute to the experimental statistics on earnings if they have a record on the Employee File. The small numbers of employees who have valid jobs on the Job File (and therefore are conceptually employees) but do not have a corresponding record on the Employee File, are excluded from the experimental statistics on earnings (although they are included in the jobs statistics).

Components of earnings

44 *Salary and wages* are the main component of earnings paid to employees, and account for the majority of cash earnings paid to employees. Salary or wages includes gross income from salary and wages, commissions, bonuses, and parental leave pay. Foreign employment earnings where tax was withheld is included in this variable as it cannot be separated from the reported salary or wages.

45 *Allowances, earnings, tips, directors fees etc.* include all employment related allowances, as well as payments from which tax was not withheld (such as tips, commissions, and bonuses).

46 *Employer lump sum payments* are payments for unused annual leave or unused long service leave.

47 *Attributed personal services income (PSI)* is a type of income generated through supply of service. Supply of service includes the skills, knowledge, expertise or efforts of an employee who performed the service on behalf of an organisation they work for. PSI is included in the concept of earnings as it can be regular, cash in nature, and attributed to employees.

48 *Employee share schemes, total assessable discount amount* is the discount on employee share scheme interests that an employee receives from their employer under an employee share scheme (ESS). The ESS discounts are treated as earnings and can be offered to the employee at any point during the financial year.

49 *Total reportable fringe benefits amount* is the gross value of all fringe benefits (in cash and in kind) reported on an ITR. This is reported only where the gross value is over \$3,738 for the financial year. Total fringe benefits valued at \$3,738 or less are not reported, which may result in slight under coverage in the experimental statistics on earnings. It is not possible to separate the cash components of fringe benefits (i.e. salary sacrifice) based on the PIT data, and thus ‘cash earnings’ is not produced separately for the LEED Foundation Projects.

50 *Reportable employer superannuation contributions* are additional to the compulsory contributions by employers. Since such contributions are made under a salary sacrifice arrangement, it is included in the concepts of earnings for the LEED Foundation Projects.

51 For further information on the ITR data items referred to above, please refer to the Individual tax

return instructions 2012 on the ATO website.

Jobs

52 A job is defined as a link between an employee and a business for \$1 or more in payment as reported on an Individual PAYG summary. An employee can have multiple jobs with the same or different businesses during the financial year, and can hold two or more jobs concurrently (see Explanatory Notes, paragraphs 58-60).

53 For the purpose of the LEED Foundation Projects, jobs only contribute to the experimental statistics if they have a record on the Job File. The employees without corresponding jobs on the Job File are excluded from the experimental statistics on jobs (although they are included in the employee statistics).

54 The concept of 'job' in other ABS publications such as Job Vacancies, Australia (cat. no. 6354.0) or the feature article "Estimating Jobs in the Australian Labour Market", Labour Force, Australia (cat. no. 6202.0) differs from the experimental statistics, as in the other publications, a job can exist independently of an employee.

55 The experimental statistics provide a volume measure of the number of filled jobs throughout the 2011-12 financial year, and do not estimate the number of jobs at a point in time. Over the reference period, an employee can contribute to the experimental statistics on jobs more than once if they have:

- multiple consecutive jobs with a single or multiple employers; and/or
- multiple concurrent jobs with a single or multiple employers.

56 The experimental statistics on jobs do not provide information on full-time equivalent jobs or unfilled jobs (job vacancies). They are more closely aligned with a volume measure of filled employee jobs during the reference period.

Main job

57 Main job is defined for each employee as the job in which they received the highest gross payment amount (see Explanatory Notes, paragraph 61) as reported on an Individual PAYG summary. This differs from ABS household surveys which define main job as the job in which the most hours were usually worked, however no information on hours worked is available in the PIT data. Gross payment may be considered a more appropriate indicator of the main job an employee held during the financial year. The main job allocated to an employee may not be the job they held at the end of the financial year.

Multiple job holders

58 Multiple job holders are employees who had two or more concurrent jobs during the reference period. The multiple job holder status of an employee is determined based on the date of information in the Individual PAYG summary. If two or more jobs were held on the same day, the employee was identified as a multiple job holder. This aligns with the definition of multiple job holder used in ABS household surveys, in which employees are only considered to be multiple jobholders if, in the reference period, they held multiple jobs concurrently.

59 The number of concurrent jobs presented in the experimental statistics are those that could be identified as concurrent using the date information. Multiple job holders may have other jobs for which concurrency could not be determined using the available date information, and as such these are not included. For employees who did not have a corresponding a job record, multiple job holder status cannot be determined.

60 Any misreporting of date information provided by an employer on an Individual PAYG summary may result in errors in the experimental statistics on multiple job holders. That is, some employees

may be incorrectly identified as multiple job holders with two or more concurrent jobs, and some actual multiple job holders may be incorrectly identified as employees who had two or more consecutive jobs during the reference period.

Gross payment

61 Gross payments are the dollar amounts recorded (by businesses) on the Individual PAYG summary for each job. Gross payments are lower than earnings as the result of two key differences:

- They include only salary and wage payments and do not include other components of earnings (see Explanatory Notes, paragraphs 36-51); and
- They are provided for each job worked by an employee, rather than as an aggregate of all jobs.

Occupation in main job

62 Occupation in main job is recorded for each employee in reference to their main job only (see Explanatory Notes, paragraph 57). Employees (or their tax agents) are asked to report their 'Main Salary and Wage Occupation' on their ITR. Employees with an ITR but without an Individual PAYG summary will have an Occupation in main job, but will not have a 'main job' recorded.

63 Occupation in main job is reported by an employee (or their tax agent) in reference to their 'Main Salary or Wage' job, in most cases this will align with the job identified from the Individual PAYG summaries as their main job. If an employee considers their main job to be something other than the job with the highest gross payment amount, the occupation they report for their main job may not relate to the main job selected from the Individual PAYG summaries.

Industry

64 For jobs linked to businesses in the *non-profiled population*, industry is based on information reported to the ATO at the time of ABN registration and is not actively maintained.

65 For jobs linked to businesses in the *profiled population*, industry is based on information collected by the ABS. Industry is collected through the profiling process for each major activity in which a business operates and is recorded separately at the TAU level.

66 As part of the integration methodology, each ABN was linked to a single TAU (and therefore a single industry). This process may result in a job (and therefore the employee) being linked to a business in a different industry to that of the actual job. This methodology was designed to preserve the distribution of employees and jobs among industries for the production of aggregate statistics (see ABN to TAU mapping in Integration Methodology). Other approaches such as linking individual jobs to TAUs may provide alternative means to perform this linking in future.

Geography

67 All geographic variables are based on a person's home address at July 2014 as reported in the PIT Client Register, and are aligned to the Australian Statistical Geography Standard (cat. no. 1270.0).

68 The Client Register is a constantly evolving register which is updated using information from various sources including ITRs. As a result, the geographic information is referenced to July 2014 when the PIT data was extracted for the LEED Foundation Projects.

69 The geographic variables for the experimental statistics therefore may not align with an employee's actual home address at the end of the reference period, 30 June 2012.

CONFIDENTIALITY

70 Taxation data is supplied to the Australian Statistician under the **Taxation Administration Act 1953** for the purposes of administering the **Census and Statistics Act 1905**. The Australian Business Register (ABR) data is supplied to the Australian Statistician under the **A New Tax System (Australian Business Number) Act 1999**.

71 In accordance with the **Census and Statistics Act 1905**, the ABS is obligated to maintain the confidentiality of individuals and businesses in these ATO and ABR data sets, as well as comply with provisions that govern the use and release of this information, including the **Privacy Act 1988**. All published statistics are subjected to a confidentiality process before release. This process is undertaken to minimise the risk of identifying particular individuals or businesses in aggregate statistics, through analysis of published data.

72 The aggregate experimental statistics on employee and job counts presented are rounded to the nearest 100.

73 The aggregate experimental statistics on earnings and gross payments are based on information as reported on the ITR or the Individual PAYG summary.

COHERENCE WITH OTHER ABS DATA

74 The experimental statistics from the LEED Foundation Projects are the result of integrating data from administrative sources. They will differ from the estimates produced from other ABS household and business collections for the following key reasons:

- There are differences in the concepts, scope, and methodology used in the LEED Foundation Projects and those used in both household and business surveys;
- The LEED Foundation Projects contain a combination of administrative data collected for taxation purposes from both individuals and businesses, and register data collected by the ABS for statistical purposes, whereas other ABS data sources are compiled for the explicit purpose of creating statistics;
- There may be some differences observed due to cash in hand and other work which is not reported to the ATO through ITR. These payments are excluded from the experimental statistics but may be included in household and business surveys if reported in the reference period; and
- The experimental statistics categorise individuals as employees if they have worked at any point during the 2011-12 financial year, whereas any point in time measure of employees includes only those who were employed during reference period.

75 The experimental statistics will differ to estimates produced from ABS household surveys which employ an 'any responsible adult' (ARA) collection methodology. The LEED Foundation Projects include information reported by individuals themselves, by their tax agent, or their employer.

76 The experimental statistics will differ from estimates produced from ABS business surveys for the following key reasons:

- ABS business surveys generally exclude employees in the Agriculture, forestry and fishing industry division, and private households employing staff. The LEED Foundation Projects include all persons in Australia who have been employed at any point during the 2011-12 financial year across all industry divisions as reported on an ITR or Individual PAYG summary;
- Many business surveys exclude irregular payments in estimates of earnings (although some may still be included if the payment was made in the reference period). These payments are included in the experimental statistics if made in the 2011-12 financial year;
- Business surveys code occupation based on the response received from the employer. The experimental statistics uses information from the ITR which is reported either by the employee or their tax agent. This may result in differences in the identification of

occupation;

- Business surveys code industry based on the ABSBR, which is based on the ABR. The experimental statistics use information about the business from the EABLD extract, which is also based on the ABSBR. For businesses in the profiled population, the ABN to TAU mapping process maintained the integrity of business-level items at the aggregate level, however, some businesses and jobs (and therefore employees) may be allocated to industries in which they do not operate. Other approaches such as linking individual jobs to TAUs may improve this allocation in future;
- The businesses included on the Integrated Dataset are selected on the basis of having a corresponding job on the Job File. They are therefore not reflective of 'active businesses in the market sector' or 'employing businesses' as used in some ABS publications such as Employee Earnings and Hours, Australia (cat. no. 6306.0) or Counts of Australian Businesses, Entries and Exits (cat. no. 8165.0); and
- For businesses in the profiled population, the ABN to TAU mapping process may result in a different distribution of business variables (e.g. Industry and Employment Size) than that in other ABS publications.

77 The experimental statistics may differ from estimates produced from ABS publications utilising PIT data due to difference in scope and due to the integration procedures used in the LEED Foundation Projects.

ACKNOWLEDGEMENTS

78 The ABS acknowledges the support provided by the ATO for the LEED Foundation Projects, and the partnership of the Department of Industry, Innovation, and Science in developing the EABLD. The provision of data as well as ongoing assistance provided by our stakeholders is essential to enable this important work to be undertaken. The enhancing of labour statistics through data integration by the ABS would not be possible without their cooperation and support.

FURTHER INFORMATION

79 For further information about the Information Paper or the experimental statistics, please contact the Labour and Income Branch on (02) 6252 7206 or via labour.statistics@abs.gov.au. The ABS Privacy Policy outlines how the ABS will handle any personal information that you provide to us.

Glossary

GLOSSARY

Age

Age of employee as at 30 June 2012 as reported on the Individual Tax Return (ITR).

Business File

Part of the Integrated Dataset. The Business File contains data from the Expanded Analytical Business Longitudinal Database extract for businesses which could be linked to a job on the Job File.

Business Turnover

The total revenue generated by a business from the provision of goods and/or services for a given accounting period (annual).

Client Dataset

This dataset contains detailed information about earnings, main occupation, tax withheld, deductions, and other items related to a single employee. This dataset is populated using

information lodged through Individual Tax Returns (ITR) to the Australian Taxation Office (ATO).

Client Register

The register of client details retained by the Australian Taxation Office (ATO), it is updated using information from the Individual Tax Returns.

Earnings

Earnings are the gross amounts paid to employees for work done or time worked (including paid leave). Earnings is the aggregate of total payments (in cash and in kind) received by each employee in all of their jobs, as reported on an Individual Tax Return (ITR).

Employee

Employees are defined as persons who worked for a private or public sector employer and received pay for the reference period in the form of wages or salaries, a commission while also receiving a retainer, tips, piece rates or payments in kind. Persons who operated their own incorporated enterprises with or without hiring employees are also included as employees.

Employee File

Part of the Integrated Dataset. The Employee File contains data relating to each employee from the Personal Income Tax (PIT) Client Register and Client Dataset.

Enterprise Group (EG)

A statistical unit covering all the operations in Australia of legal entities under common control. Multiple Australian Business Numbers (ABNs) can operate within a single Enterprise Group (EG), and each EG is broken up into one or more Type of Activity Units (TAUs).

Employment size

The number of employees in a business. This is updated annually (for non-profiled businesses) or reported at a point in time during the profiling process (profiled businesses).

First job and second job

First job and second job refer to jobs held by a multiple job holder (see Multiple job holders). First job is the job with the highest gross payment reported on an employee's Individual Pay As You Go (PAYG) summaries which is also held concurrently with another job. First job may differ from the employee's main job. Of the jobs with which first job is held concurrently, the job with the second highest gross payment is selected as the second job.

Geography

All geographic variables are based on an employee's home address at July 2014 as reported in the Personal Income Tax (PIT) Client Register, and are aligned to the Australian Statistical Geography Standard (ACGS): Volume 1 – Main Structure and Greater Capital City Statistical Areas, July 2011 (cat. no. 1270.0.55.001). For further information see Explanatory Notes (paragraphs 67-69).

Gross payment

Gross payments are the dollar amounts recorded (by businesses) on the Individual Pay As You Go (PAYG) summary for each job.

Individual PAYG summary

The annual summary provided by an employer to the Australian Taxation Office (ATO) with respect to an employee. It records job level information reported by employers about the gross payment made to an employee, tax withheld, and the start and end dates for each job. This also provides the Australian Business Number (ABN) of the employer. This usually has a Tax File Number (TFN) attached, although in some circumstances this may be missing or substituted for another code (e.g. if the employee did not provide it or is under the age of 18 and earns less than the tax-free threshold).

Individual Tax Return (ITR)

The annual tax return submitted by individuals to the Australian Taxation Office (ATO).

Industry (ANZSIC)

Industry information of each employing business aligns with the Australian and New Zealand Standard Industrial Classification (ANZSIC), 2006 (cat. no. 1292.0). The structure of ANZSIC comprises four levels, ranging from industry division (broadest level) to industry class (finest level). In this release, industry is provided at the division level. The industry division provides a limited number of categories which give a broad overall picture of the economy. There are 19 divisions within ANZSIC, each identified by an alphabetical letter, that is, 'A' for Agriculture, forestry and fishing, 'B' for Mining, 'C' for Manufacturing, etc.

Institutional sector (SISCA)

Institutional sector of each employing business. This is aligned with the Standard Institutional Sector Classification of Australia outlined in Standard Economic Sector Classifications of Australia (SESCA), 2008 (cat. no. 1218.0). In this release, Institutional sector is provided at the sector level.

Integrated Dataset

This is the physical file which constitutes the linked employer-employee data. The Integrated Dataset is comprised of three main subsets:

- Employee File
- Job File
- Business File.

Job

A job is defined as a link between an employee and a business for \$1 or more in payment as reported on an Individual Pay As You Go (PAYG) summary. An employee can have multiple jobs with the same or different businesses during the financial year, and can hold two or more jobs concurrently.

Job File

Part of the Integrated Dataset. The Job File contains data relating to each job from the PIT Individual Pay As You Go (PAYG) Dataset.

Main job

Main job is defined for each employee as the job in which they received the highest gross payment amount as reported on an Individual Pay As You Go (PAYG) summary.

Multiple job holders

Multiple job holders are employees who had two or more concurrent jobs during the reference period. The multiple job holder status of an employee is determined based on the date of information in the Individual PAYG summary. If two or more jobs were held on the same day, the employee was identified as a multiple job holder.

Occupation in main job

An occupation is a collection of jobs that are sufficiently similar in their title and tasks, skill level and skill specialisation which are grouped together for the purposes of classification. Occupation refers to Major Group as defined by the Australian and New Zealand Standard Classification of Occupations, First Edition, Revision 1, Revision 1 (cat. no. 1292.0) of the job which the employee identifies as their 'Main Wage or Salary Job'.

Public and private sector

This aligns with the Type of Legal Organisation outlined in Standard Economic Sector Classifications of Australia (SESCA), 2008 (cat. no. 1218.0). See Type of Legal Organisation (TOLO).

Statistical Area Level 4 (SA4)

An area defined in the Australian Statistical Geography Standard and designed for the output of labour force data and to reflect labour markets. In rural areas, SA4s generally represent

aggregations of multiple small labour markets with socioeconomic connections or similar industry characteristics. Large regional city labour markets are generally defined by a single SA4. Within major metropolitan labour markets SA4s represent sub-labour markets. SA4s are built from whole Statistical Area Level 3s. They generally have a population over 100,000 people to enable accurate labour force survey data to be generated. There are 88 SA4s and they cover the whole of Australia without gaps or overlaps.

For further information, refer to Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas (cat. no. 1270.0.55.001).

Sex

Sex of employee as at 30 June 2012 as reported on the Individual Tax Return (ITR).

Start and End Dates

Start and end dates associated with each job as reported on an Individual Pay As You Go (PAYG) summary.

State and territory

See Geography.

Type of Activity Unit (TAU)

The TAU, residing in the profiled population, comprises one or more business entities, sub entities or branches of a business entity within an enterprise group (EG). These entities can report production and employment data for similar economic activities when a minimum set of data items are available. TAU is the statistical unit used to define a business in the profiled population in the Linked Employer-Employee Database Foundation Projects.

Type of Legal Organisation (TOLO)

The type of legal organisation of each employing business. This is aligned with the Type of Legal Organisation classification outlined in Standard Economic Sector Classifications of Australia (SESCA), 2008 (cat. no. 1218.0). In this release, Type of Legal Organisation is provided at the group level.

Abbreviations

ABBREVIATIONS

ABN	Australian Business Number
ABR	Australian Business Register
ABS	Australian Bureau of Statistics
ABSBR	Australian Bureau of Statistics Business Register
ANZSIC	Australian and New Zealand Standard Industrial Classification
ATO	Australian Taxation Office
AWE	Average Weekly Earnings
EABLD	Expanded Analytical Business Longitudinal Database
EEBTUM	Employee Earnings, Benefits and Trade Union Membership
EEH	Employee Earnings and Hours
EG	Enterprise Group
ITR	Individual Tax Return
LEED	Linked Employer-Employee Database
LFS	Labour Force Survey
OMIEs	Owner Managers of Incorporated Enterprises
OMUEs	Owner Managers of Unincorporated Enterprises
PAYG	Pay As You Go

PIT	Personal Income Tax
SA 4	Statistical Area level 4
SISCA	Standard Institutional Sector Classification of Australia
STFN	Scrambled Tax File Number
TAU	Type of Activity Unit
TFN	Tax File Number
TOLO	Type of Legal Organisation

Appendix 1: Distribution of Key Variables (Appendix)

APPENDIX 1: DISTRIBUTION OF KEY VARIABLES

To assist users in interpreting the aggregate experimental statistics, this appendix presents the distribution of key data items on the Integrated Dataset, including the amount of inadequately defined, not stated or missing information on each of the three files.

EMPLOYEE FILE AND JOB FILE

The following section presents the distribution of data from the Employee File and Job File. Care should be taken in comparing these to other publications, as the population and the distribution of variables will be different (see Explanatory Notes, paragraphs 74-77).

Age

Table 1.1: Distribution of employees by age group

Age group	(‘000)	Employees %
14 years and under	5.7	0.1
15-17	134.8	1.3
18-20	561.0	5.4
21-24	992.2	9.6
25-29	1 337.5	12.9
30-34	1 192.8	11.5
35-39	1 113.5	10.8
40-44	1 142.2	11.1
45-49	1 071.2	10.4
50-54	1 029.9	10.0
55-59	826.2	8.0
60-64	564.4	5.5
65-69	238.2	2.3
70-74	71.4	0.7
75-79	27.5	0.3
80-84	15.8	0.2
85 & over	10.3	0.1

Sex

Table 1.2: Distribution of employees by sex

Sex	(‘000)	Employees %
Male	5 359.1	51.9
Female	4 975.6	48.1

Geography

Table 1.3: Distribution of employees by state and territory of home address

State and territory	('000)	Employees %
New South Wales	3 180.7	30.8
Victoria	2 531.1	24.5
Queensland	2 065.7	20.0
South Australia	705.9	6.8
Western Australia	1 151.5	11.1
Tasmania	216.7	2.1
Northern Territory	95.0	0.9
Australian Capital Territory	198.3	1.9
Other territories	0.5	0.0
Inadequately defined/missing	189.3	1.8

Table 1.4: Completeness of address coding at the state and territory and SA4

Geographic level	('000)	Employees %
State and territory present	10 145.4	98.2
State and territory missing	189.3	1.8
SA4 present	10 123.0	98.0
SA4 missing	211.7	2.1

Occupation in Main Job**Table 1.5: Distribution of employees by occupation (a)**

Occupation	('000)	Employees %
Managers	1 213.5	11.7
Professionals	2 180.3	21.1
Technicians and trades workers	1 274.0	12.3
Community and personal service workers	1 037.9	10.0
Clerical and administrative workers	1 624.7	15.7
Sales workers	892.3	8.6
Machinery operators and drivers	624.2	6.0
Labourers	1 062.7	10.3
Inadequately defined/not stated	425.2	4.1

(a) Occupation in main job.

Earnings and Gross Payments

The following table shows the total earnings (reported on an ITR) for all employees, and gross payments (reported on the Individual PAYG summaries) for all jobs, along with their respective mean and median.

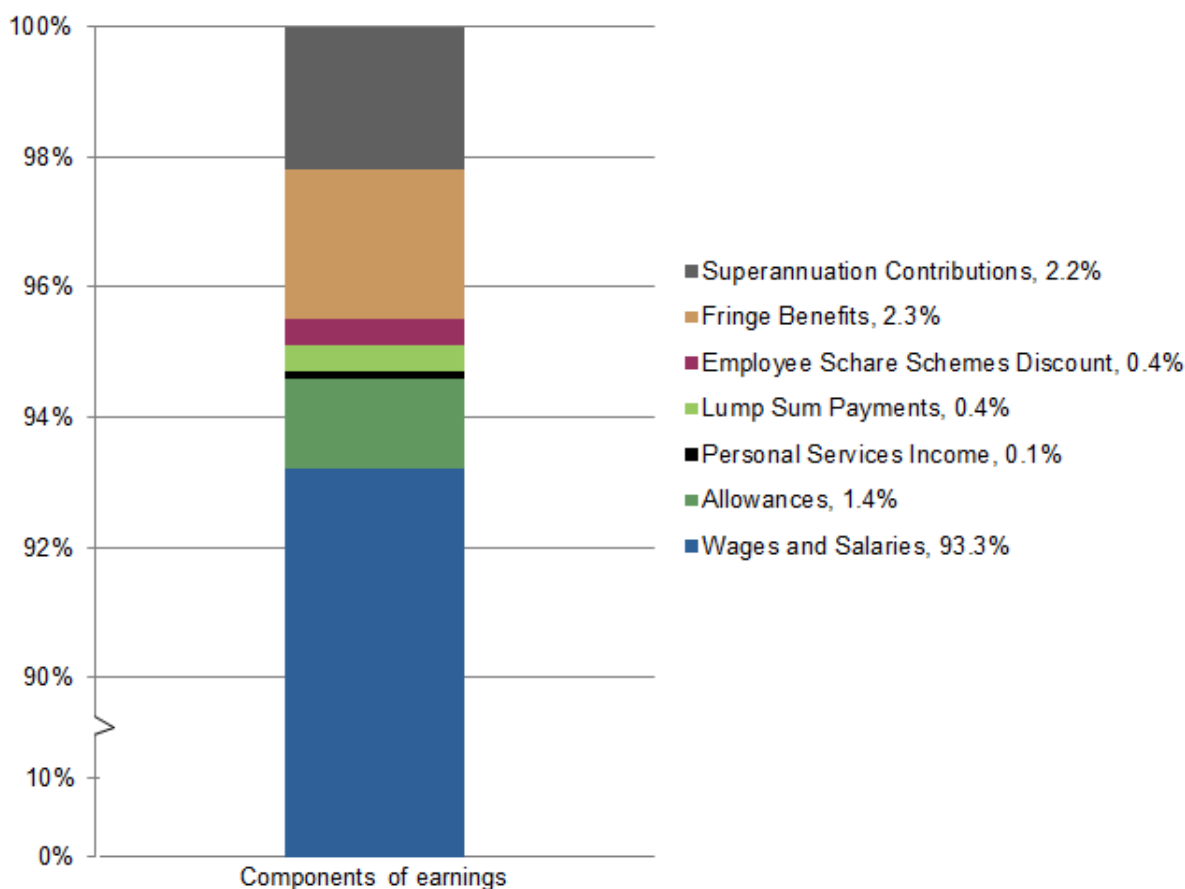
Table 1.6: Earnings and gross payments in all jobs

Item	Total (\$b)	Mean (\$)	Median (\$)
Earnings	575.4	55 678	45 869
Gross payments	505.5	37 961	26 134

Mean earnings is calculated by summing all employee earnings (Total) and dividing by the number of employees. Mean gross payment is calculated by summing all of the gross payments (Total) and dividing by the number of jobs.

The following graph shows the proportional distribution of the components of earnings.

Graph 1.1: Distribution of earnings components (a)



(a) This excludes employees whose earnings data was cleaned (see Integration Methodology).

As shown in Graph 1.1, Wages and Salaries comprise 93.3% of earnings, with the remaining 6.7% distributed mostly among Superannuation Contributions, Fringe Benefits, and Allowances.

The distribution of earnings recorded in the LEED Foundation Projects is comparable at aggregate level to the data presented in the ATO Tax Statistics for the 2011-12 financial year published on the ATO website.

Start and End Dates

The following table shows the quality of the start and end dates for each job. It is categorised by whether one or both dates are present and whether they are within an acceptable range (between 1 July 2011 to 30 June 2012).

Table 1.7: Comparison of dates from Individual PAYG summary data

Quality of date variable	('000)	Jobs %
Both dates acceptable	13 092.0	98.3%
Date missing		
One date	10.6	0.1%
Both dates	73.9	0.6%
Error in dates		
One date	116.7	0.9%
Both dates (a)	23.7	0.2%

(a) Includes records where the end date is before the start date.

As shown in Table 1.7, the start and end dates for some Individual PAYG summaries contain invalid date information. These resulted in durations of 0 or negative days, or more than 366 days. The number of incorrect and missing dates, however, is relatively low at 1.7%. There may be other incorrect dates which fall within the 2011-12 financial year but these are not able to be identified.

BUSINESS FILE

The following section presents the distribution of data from the Business File. Care should be taken in comparing these to other ABS publications such as Counts of Australian Businesses, Entries and Exits (cat. no. 8165.0), as the population and the distribution of variables will be different as a result of the scope (see Introduction) and the ABN to TAU mapping process (see Integration Methodology) used in the LEED Foundation Projects (see Explanatory Notes, paragraphs 74-77).

Profiled and Non-profiled businesses

Table 1.8: Distribution of businesses by profiled status

Whether profiled	('000)	Businesses %	Job share (a) %
Profiled (TAU based)	7.8	1.1	49.2
Non-profiled (ABN based)	675.6	98.9	50.7

(a) Percentage of all jobs on the Job File. Totals do not add to 100% due to jobs which are not linked to businesses.

Employment size

Table 1.9: Distribution of businesses by employment size

Employment size	('000)	Businesses %	Employment share (a) %
1 to 4 employees	419	61.3	9.4
5-19 employees	172	25.2	16.0
20-199 employees	48.8	7.1	23.5
200 employees or more	4.2	0.6	51.1
No employees reported (b)	39.3	5.8	0.0

(a) Percentage of total employment on the Business File.

(b) These businesses have jobs linked to them in the LEED Foundation Projects, but have reported no employees. This may be due to:

- the time of reporting (i.e. they did not have any employees at the time of reporting but did at another point in the financial year); or
- incorrect linking of jobs to a TAU; or
- reporting errors made by the business.

Business Turnover

Table 1.10: Distribution of businesses by business turnover

Business turnover	('000)	Businesses %	Turnover share (a) %
Less than \$50,000	43.5	6.4	0.0
\$50,000 to less than \$200,000	144.9	21.2	0.6
\$200,000 to less than \$2 million	376.6	55.1	8.1
\$2 million or more	100.5	14.7	91.3
Inadequately defined/not stated	17.7	2.6	0.0

(a) Percentage of turnover on the Business File.

Industry

Table 1.11: Distribution of businesses by industry

Industry	(‘000)	Businesses %
Agriculture, forestry and fishing	46.0	6.7
Mining	3.1	0.5
Manufacturing	41.1	6.0
Electricity, gas, water and waste services	2.1	0.3
Construction	104.1	15.2
Wholesale trade	33.2	4.9
Retail trade	64.4	9.4
Accommodation and food services	46.3	6.8
Transport, postal and warehousing	32.1	4.7
Information media and telecommunications	5.8	0.8
Financial and insurance services	24.7	3.6
Rental, hiring and real estate services	25.3	3.7
Professional, scientific and technical services	91.4	13.4
Administrative and support services	29.7	4.4
Public administration and safety	3.9	0.6
Education and training	13.3	2.0
Health care and social assistance	47.1	6.9
Arts and recreation services	9.6	1.4
Other services	52.0	7.6
Inadequately defined/not stated	8.0	1.2

Type of Legal Organisation (TOLO)

Table 1.12: Distribution of businesses by TOLO

TOLO	(‘000)	Businesses %
Incorporated private sector entities	377.1	55.2
Unincorporated private sector entities	294.4	43.1
Public sector entities	11.8	1.7

Institutional Sector (SISCA)

Table 1.13: Distribution of businesses by institutional sector

Institutional sector	(‘000)	Businesses %
Non-financial corporations	351.0	51.4
Financial corporations	24.5	3.6
General government	3.3	0.5
Households	281.3	41.2
Non-profit institutions serving households	15.1	2.2
Inadequately defined/not stated	8.1	1.2

Appendix 2: Coherence of Experimental Statistics with ABS Survey Collections (Appendix)

APPENDIX 2: COHERENCE OF EXPERIMENTAL STATISTICS WITH ABS SURVEY COLLECTIONS

The data used in the following comparisons is compiled from a variety of ABS survey collections. The Labour Force Survey (LFS) data is based on an average of the quarterly original data (see Labour Force, Australia, Detailed, Quarterly (cat. no. 6291.0.55.001)) for the 2011-12 financial year. The Average Weekly Earnings (AWE) data is based on averages of quarterly trend data (see Average Weekly Earnings (cat. no. 6302.0)) for all four quarters during the 2011-12 financial year. The Employee Earnings and Hours (EEH), Australia, (cat. no. 6306.0) data is for the May 2012 reference period, and the Employee Earnings, Benefits and Trade Union Membership (EEBTUM), Australia (cat. no. 6310.0) is for the August 2011 reference period. The experimental statistics on mean weekly earnings are calculated by dividing the annual earnings by 52.29 weeks (this differs slightly from the usual 52.14 weeks as 2012 was a leap year).

GENERAL DIFFERENCES

The general differences in the number of employees and earnings were due to:

- differences in the concepts, scope and methodology used in the LEED Foundation Projects and those used in household and business surveys;
- the LEED Foundation Projects containing a combination administrative data collected for taxation purposes from both individuals and businesses, and ABS Business Register data collected for statistical purposes, whereas other ABS data sources are compiled for the explicit purpose of creating statistics;
- unreported cash in hand payments which are excluded from the experimental statistics but may be included in household and business surveys if reported in the reference period; and
- the experimental statistics categorising individuals as employees if they had worked at any point during the 2011-12 financial year, whereas any point in time measure of employees includes only those who were employed during reference period (often the last week/fortnight or last pay period on or before a specified cut-off date).

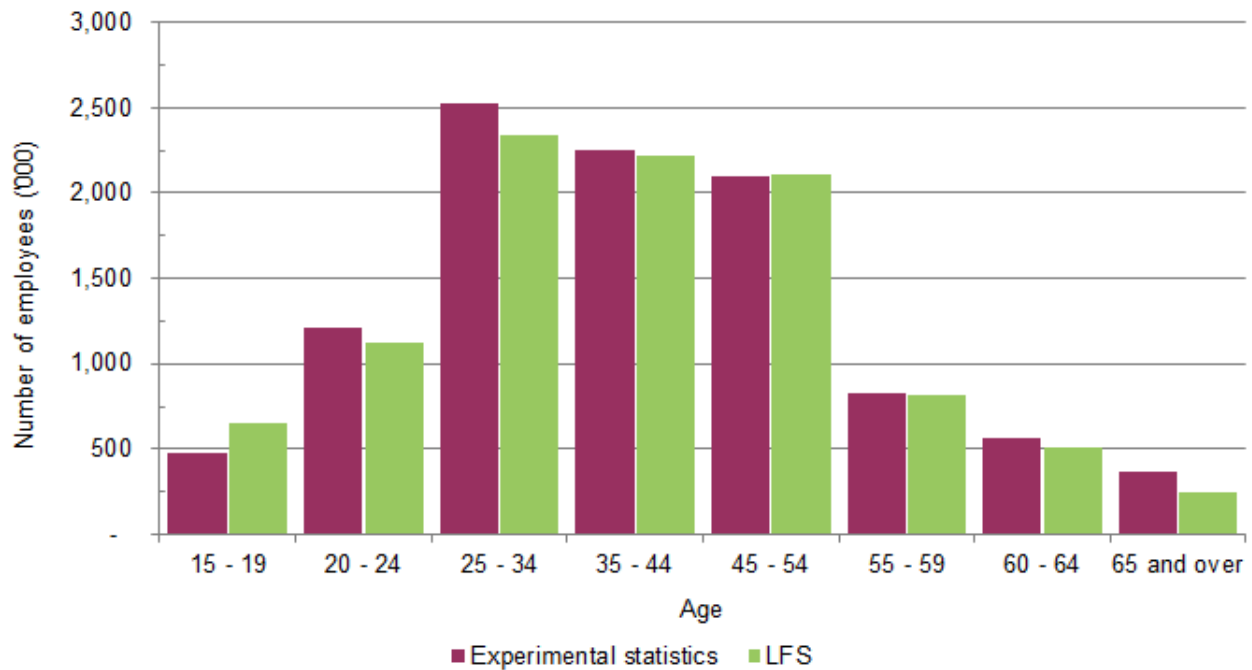
NUMBER OF EMPLOYEES

The following graphs show comparisons between the aggregate experimental statistics and estimates of the number of employees from selected ABS business and household surveys.

Age

There were minimal differences found in the number of employees by age group between the experimental statistics and LFS, especially for employees aged between 35 and 59 years. The experimental statistics are expected to be higher for younger and older employees due to more frequent periods without employment during the financial year (observed in the 20 to 34 year age groups in particular, as well as the 60 and over age groups due to bridge retirement practices). The experimental statistics are expected to be lower for younger employees due to unreported cash in hand work (observed in the 15 to 19 age group).

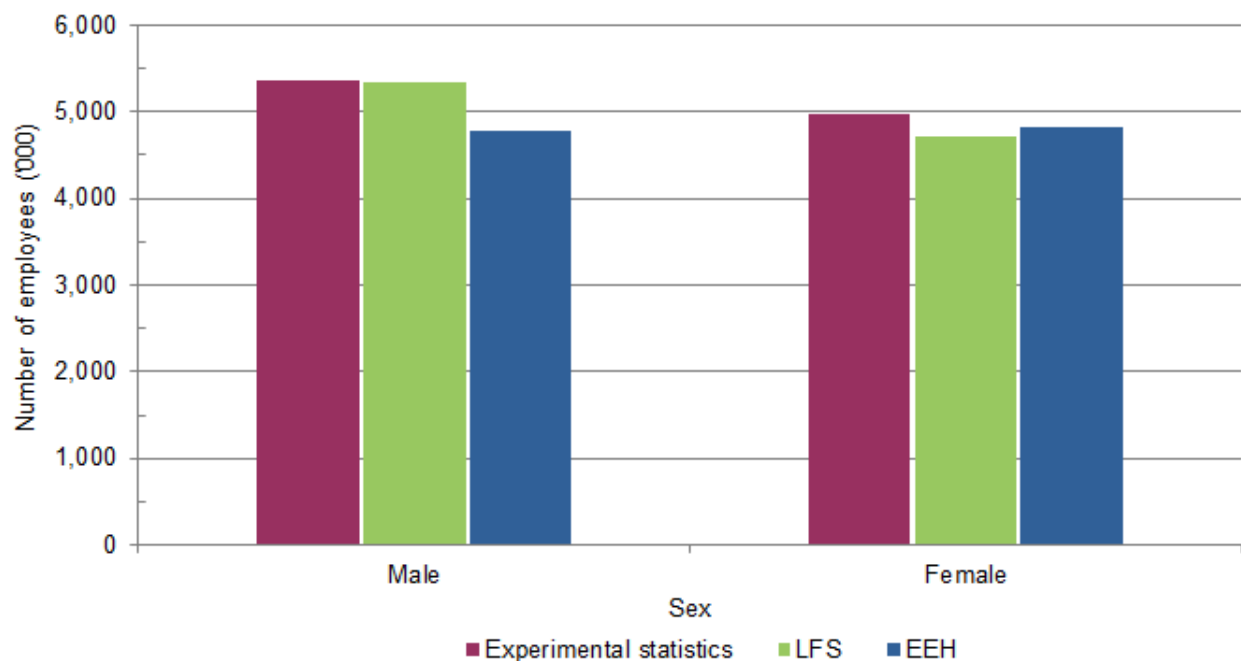
Graph 2.1: Number of employees, by age, 2011-12



By sex

Due to the way employees are defined and measured in the LEED Foundation Projects, it was expected that the experimental statistics would be more coherent with estimates of employment from household rather than business surveys. However, the LEED Foundation Projects measure the gross volume of employees in the financial year, which would inflate the experimental statistics against all sources, while employment not reported to the ATO will deflate them. For male employees, there is no statistically significant difference between the experimental statistics and the LFS estimates, whereas for females there is a 6% difference. This may be due to greater numbers of women transitioning between labour force status during the financial year. The difference between the experimental statistics and EEH is much higher for males (12%) and for females (3%). The differences are likely caused by general differences in concepts, scope and methodology employed in EEH. The exclusion of the Agriculture industry (which employs approximately twice as many males as females) from EEH may deflate the number of male employees in EEH compared to the experimental statistics.

Graph 2.2: Number of employees, by sex, 2011-12



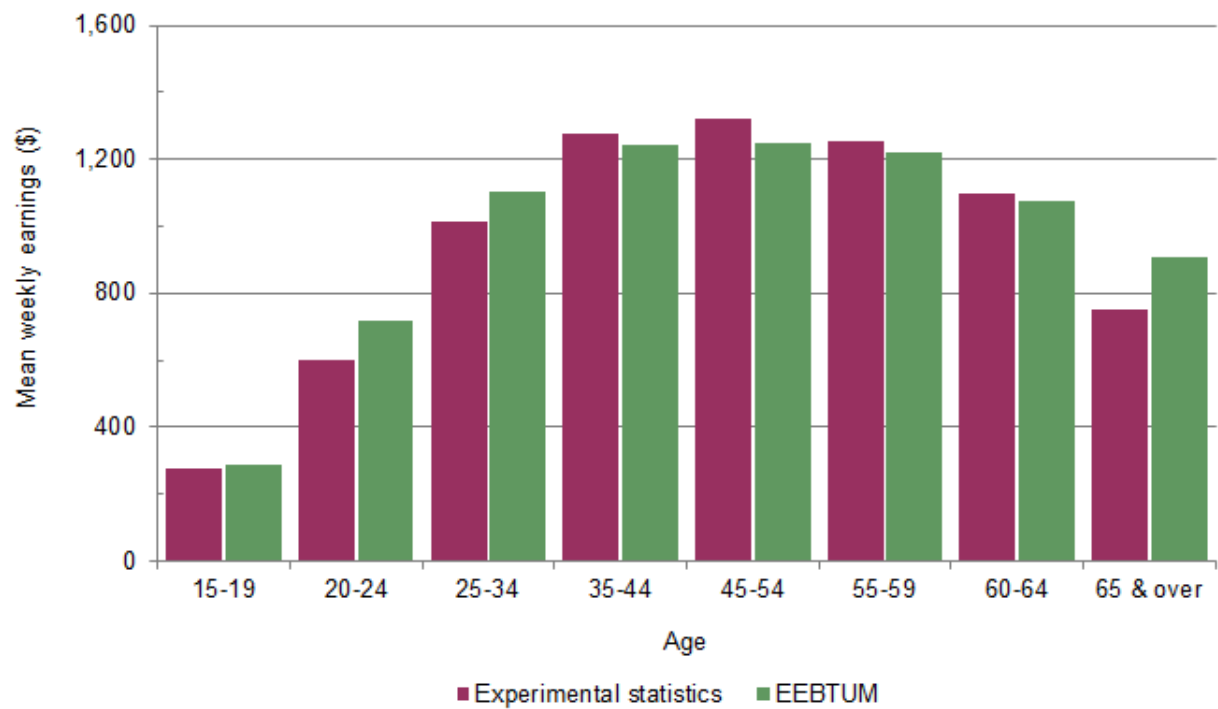
EMPLOYEE EARNINGS

The following graphs show comparisons of average weekly earnings between the experimental statistics and the estimates from selected ABS business and household surveys.

Age

It is expected that average weekly earnings from the LEED Foundation Projects will be broadly coherent with EEBTUM. Periods without employment during the financial year (more prominent for younger or older employees) as well as unreported cash in hand work (also more common in younger age groups) are expected to deflate the experimental statistics (as observed in the 15 to 34 year age groups, as well as the 65 and over age group). The differences in the 65 and over age group were potentially influenced by the tendency of some older employees not to report their earnings on an ITR. The experimental statistics include the gross value of fringe benefits, while EEBTUM includes only the salary sacrificed component, which will inflate the experimental statistics, particularly in the older age groups where fringe benefits are more prominent. There is no statistically significant difference in the 60 to 64 year age group, which is potentially due to a balancing of these effects.

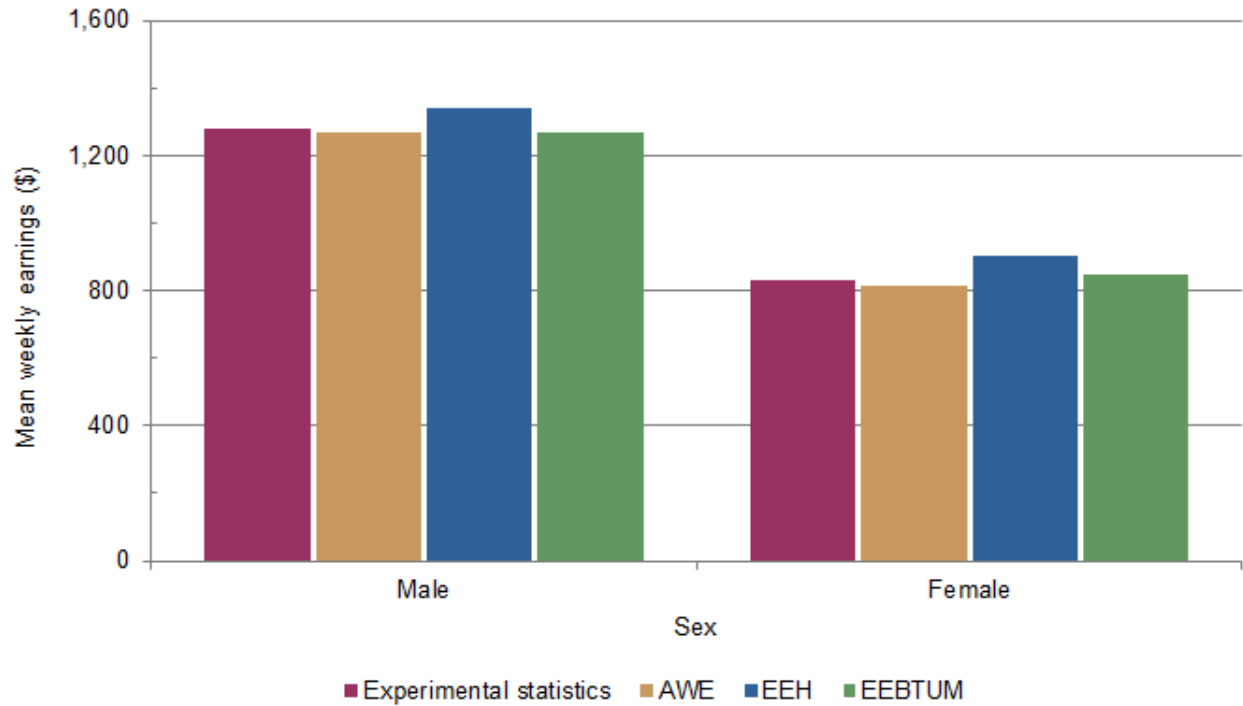
Graph 2.3: Mean weekly earnings, by age, 2011-12



Sex

There is no statistically significant difference between mean weekly earnings from the experimental statistics and estimates from AWE. There are minimal differences between the experimental statistics and EEBTUM, which are likely due to the inclusion of fringe benefits (inflating the experimental statistics) and unreported cash in hand work (deflating the experimental statistics), as well as seasonal variability in the EEBTUM estimates. The EEH estimates are higher (4.7% for males and 8.6% for females) than the experimental statistics (as well as AWE and EEBTUM). This is likely due to the general differences in concepts, scope and methodology employed in EEH, including the exclusion of Agriculture (which has lower mean earnings).

Graph 2.4: Mean weekly earnings, by sex, 2011-12



© Commonwealth of Australia

All data and other material produced by the Australian Bureau of Statistics (ABS) constitutes Commonwealth copyright administered by the ABS. The ABS reserves the right to set out the terms and conditions for the use of such material. Unless otherwise noted, all material on this website – except the ABS logo, the Commonwealth Coat of Arms, and any material protected by a trade mark – is licensed under a Creative Commons Attribution 2.5 Australia licence